Automation in Construction 132 (2021) 103940

Contents lists available at ScienceDirect

Automation in Construction

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

Review

Computer vision applications in construction: Current state, opportunities & challenges

Suman Paneru, Idris Jeelani *

Rinker School of Construction Management, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Keywords:
Computer vision
Technology in construction
Visual data analytics
Monitoring

ABSTRACT

Thousands of images and videos are collected from construction projects during construction. These contain valuable data that, if harnessed efficiently, can help automate or at least reduce human effort in diverse construction management activities such as progress monitoring, safety management, quality control and productivity tracking. Extracting meaningful information from images requires the development of technology and algorithms that enable computers to understand digital images or videos, replicating the functionality of human visual systems. This is the goal of computer vision. This review aims at providing an updated and categorized overview of computer vision applications in construction by examining the recent developments in the field and identifying the opportunities and challenges that future research needs to address to fully leverage the potential benefits of Computer Vision. We restrict the focus to four areas that can benefit the most from computer vision - Safety Management, Progress Monitoring, Productivity Tracking and Quality Control.

1. Introduction

An image is said to be worth 1000 words. More than 400,000 images are captured from a typical construction project (∼17,000 Sft.) during its construction phase [1]. This number is rising with the advancement of technology and the ever-increasing use of camera-equipped devices such as drones, ground robots, smartphones, and tablets on construction sites. This enormous volume of images and videos contains a treasure of valuable data, which can potentially be harnessed for a variety of project management activities such as surveillance, progress monitoring, safety management, quality inspections, resource utilization management and others. Traditionally, images and videos captured on construction sites have been used for documenting and tracking the status of the project [2], documenting the safety and quality inspections [3], keeping a visual timeline of site progress, providing evidence against damage claims [4], capturing workmanship, and providing field updates to the office. In addition, construction projects use surveillance cameras on the jobsite boundary for security purposes [5]. In most cases, the images and videos are manually examined, and their use has been largely been limited to documentation and record-keeping [6,7]. More importantly, only a limited amount of information is extracted and used from these images. For example, the images collected by surveillance cameras are predominantly used to record any unauthorized intrusion to the site [5]. However, the visual data captured by surveillance cameras contains other useful information as well, such as the progress of work over time [8], safety compliance by workers [9], idle time of equipment, material usage and much more, which remains largely unused. To prevent this underutilization of data, there is a need for efficient technology to automatically extract and analyze valuable and meaningful information captured in images and videos. In recent years, there has been a paradigm shift in construction and the role of visual data is changing from being a passive instrument of documentation and record-keeping to an active tool in project management. For example, images and videos are used to detect defects and assess conditions of concrete and asphalt in civil infrastructures [10,11]. Similarly, data from images are processed to automatically detect whether or not workers are wearing their hard hats [12,13]. Such automated systems help construction managers make informed decisions for efficient safety, productivity, and quality management. However, automated extraction of information from images and videos is a challenging task [14,15] and requires the development of technology that enables a computer to understand an image by detecting, identifying, and classifying various objects present in an image, just like a human vision system. This is the goal of computer vision (CV), which is an interdisciplinary field aimed at developing algorithms to enable computers to understand digital images or videos, replicating the functionality of human visual systems [16,17]. A typical computer

* Corresponding author.
E-mail addresses: spaneru@ufl.edu (S. Paneru), idris.jeelani@ufl.edu (I. Jeelani).

https://doi.org/10.1016/j.autcon.2021.103940
Received 5 April 2021; Received in revised form 30 August 2021; Accepted 1 September 2021
Available online 15 September 2021

vision system accepts 2D images (or videos) as an input, transforms it into a mathematical form (using pixel values), analyzes this data to recognize meaningful patterns, distinctive features, spatial arrangement, among other things, and provide a description of the image as detailed as required by the problem [18]. Fig. 1 shows a typical pipeline of a computer vision-based system.

Computer vision techniques are widely used in todays' world in almost every field. The most common example is perhaps the face detection ability of our smartphone cameras. Popular self-driving cars also use various vision-based techniques such as Simultaneous Localization and Mapping (SLAM) and object recognition to make decisions while driving autonomously on roads [16]. The other examples include the detection of cancer cells from CT and PET scans [19], detection and classification of skin lesion [19] in medicine, prediction of traffic speed in transportation, quality inspection of packaging [20], identification of defective products and remote inspections of pipelines and equipment in manufacturing [21]. In the mars exploration mission, researchers used stereo vision and visual odometry for rover navigation and feature tracking for horizontal velocity estimation of the landers [22]. In construction, computer vision is increasingly being used for safety management [10], quality inspections [23], productivity monitoring [24], and navigation of unmanned ground and aerial vehicles [6]. Some of the examples include estimation of worker's pose [25], detection of PPE [26,27], comparing actual vs planned work [28], and conducting the quality inspection of critical infrastructure [14].

1.1. Computer vision in construction

Construction is on the path of increasing automation. Owing to the large amount of visual data it generates, the construction industry can greatly benefit from the automatic extraction and analysis of this useful data. Computer vision can help automate several construction managements tasks that currently require extensive human involvement for visual examination. This includes safety monitoring (E.g., detecting noncompliance to PPE requirement), quality inspections (E.g., detecting installation defects), progress monitoring (E.g., comparing as-built 3D geometry with as-planned 3D/4D model), navigation assistance (E.g., proximity alerts for construction vehicles), automated/ robotic construction (E.g., controlling robotic arms of painting or brick laying robot). Even though vision-based techniques are increasingly being used in construction, there is still significant untapped potential in this area that future research needs to explore.

1.2. Goals and objectives

This study aims at reviewing the current state of computer vision (CV) in construction from a holistic approach and identifying the opportunities and challenges that future research needs to address to fully leverage the potential benefits of CV in construction. First, we discuss the most common and important computer vision tasks relevant to various construction management applications. Second, the study aims to provide an updated overview of computer vision applications in construction that captures the recent developments in this rapidly evolving area. To maintain a homogeneous set of contributions, we restrict the focus to the use of different computer vision techniques in four areas within construction that can most benefit from computer vision. These are progress monitoring, safety monitoring, quality control and automated construction. Third, the study aims to structure the obtained information in a way that computer vision research in different areas can easily be linked to each other and compared on multiple facets, which will facilitate future research works within a specific researcher's area of interest. The paper also identifies specific challenges and future research opportunities for the integration of computer vision in construction.

1.3. Point of departure

As of January 2021, there have been a few important reviews of computer vision research in construction. [29] conducted a comprehensive review of computer vision applications in construction safety assurance. In addition, [30] conducted a scientometric review of computer vision research for construction applications and [31] mapped the computer vision research in construction in 2019. While all three works have made valuable contributions, the current work offers different contributions. For example [29] focused only on safety assurance, whereas this work takes a holistic approach to review computer vision applications for different construction management activities not limited to any one area. Secondly [30,31] predominantly focused on mapping the published work on computer vision in construction. The results provide trends of published work, author and co-other analysis to identify researchers involved in such research, geographic mapping of published work etc. In contrast, the current work focuses on the

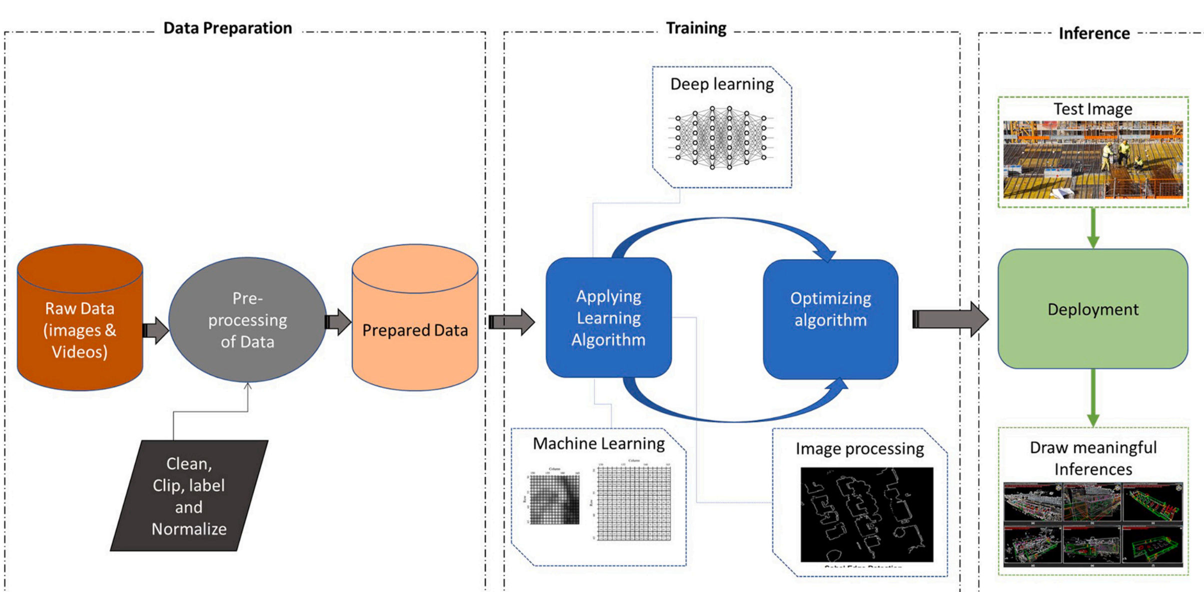

Fig. 1. Typical computer vision process.

application of different computer vision tasks in different areas of construction management. While studies [30,31] focused more on *"who"* and *"where",* the current study focuses on "*what*" and *"how"* of the computer vision research in construction. Compiling literature in this manner enables the identification of issues, opportunities and challenges that are not currently well covered and examined. This study lays foundation work for future research aimed at approaching computer vision in construction from a holistic perspective or focusing on developing computer vision applications in specific areas within construction.

1.4. Contributions

This paper provides the latest review of computer vision-based technology in construction to evaluate the current state of computer vision applications, identify challenges in implementation and future research opportunities. The paper categorizes the applications by 1) different areas of construction management (i.e. safety management, progress monitoring, quality control & productivity analysis) and 2) different computer vision tasks (such as classification, object recognition, object tracking, action recognition).

1.5. Organization

The paper is organized as follows: We start with Section 1 that discusses the goals of the review and its point of departure from similar past works. Section 2 defines and provides a concise explanation for various computer vision tasks that are commonly used in construction. Section 3 presents the review methodology and discusses the inclusion and exclusion criteria used to filter the literature. Section 4 presents the review findings, and the subsections categorize the applications by different areas of construction management. This is followed by Section 5 that highlights the challenges and opportunities of computer vision techniques in construction from the holistic perspective. Finally, Section 6 presents the conclusion.

2. Background

Computer Vision (CV) is a broad field that encompasses several distinct vision techniques accomplished for different objectives and predictions. For example, it can be used to create a digital twin of a construction (*3D Scene Reconstruction*) or used to detect hazards on construction sites (*object detection*) or used to find an obstacle-free path for an autonomous ground vehicle (*segmentation*). Modern computer vision techniques can be traced back to an ambitious MIT summer project by Seymour Papert and Marvin Minsky in 1966 [18]. The goal of this project was to build a system that can analyze images and identify objects in these images. Even though the goal was not achieved, it is often said to have laid the foundation of modern-day computer vision [18]. In 1979, Fukushima [17] proposed the Neocognitron, which included a hierarchical, multilayered artificial neural network used for handwritten character recognition and other pattern recognition tasks. It served as the inspiration for convolutional neural networks (CNN) developed in 1980. A convolutional neural network consists of multiple layers of artificial neurons, which are mathematical components similar in functioning to biological neurons [18]. Unlike previous systems, where the image was processed holistically, each layer of a CNN extracts specific features from the pixels of the images. For example, the initial layers detect basic features, such as vertical and horizontal edges, deeper layers use these simple features to detect more complex features such as corners and basic geometric shapes, and the final layers use these complex features to detect specific entities or objects of our interest such as faces, doors, and cars [19,32,33]. While CNNs performed exceptionally well compared to previous attempts, the amount of data and computational resources needed to tune and use CNNs was extremely high, limiting their use to banking and postal services only [34]. Hence, most computer vision problems used machine learning techniques such as support vector machines (SVM) and random forest [32]. These machine learning approaches utilize "features," (a measurable piece of data that is unique to this specific object such as distinct pattern, color etc.) and use a statistical learning algorithm to detect objects in images or classify images based on these features. However, identifying the features specific to objects requires enormous human and computational resources. In 2012 CNNs regained popularity with the development of ALEXNET [17,35], which demonstrated great potential. Since then, advances in deep learning have enabled computer vision to grow immensely and it has found applications in almost every field from medicine to defense, to transportation to manufacturing and of course, construction. Fig. 2 shows the evolution of computer vision in construction listing some of the notable works that represent important milestones in the path of computer vision integration in construction.

Computer vision is a broad field that encompasses several techniques used to extract and process visual data from images and videos to draw meaningful inferences. Some of these techniques that are important to construction management tasks are detailed below.

2.1. 3D scene reconstruction

3D Scene Reconstruction is a process to create 3D models of a scene from a set of 2D images. By applying the 3D scene reconstruction process, 3D shapes of complex objects can be modelled provided the 2D images contain all the required information [36]. Early phases of 3D scene reconstruction research developed a mathematical process to understand 3D to 2D conversion process to develop algorithmic solutions, which later became the foundation for the development of 3D reconstruction. In construction, 3D reconstruction is used in construction progress monitoring, structure inspection and post-disaster rescue. A typical process involves building an as-built 3D model from 2D images (or laser point clouds) captured on site. This model can be used for a variety of applications such as comparing progress over time, conducting quality inspections, inspecting mechanical structures (plumbing, electrical, HVAC systems) or visualization purposes. For example, Fig. 3 shows an interesting work by Han and Fard [28] demonstrating the use of 3D reconstruction to monitor the progress of work. As shown in the figure, a 3D model of the construction site is created by a set of 2D images obtained from the construction site and this as-built 3D model is compared with the as-planned BIM model to track progress.

As sensors, such as a camera, can only capture visible information, perceiving 3D shapes of an object, its volumetric composition, and the overall information is daunting for machines, unlike a human who can perceive visible as well as invisible information while examining an image [37]. While capturing images, the 3D geometry is projected into the 2D image sensor and as a result, the depth information is lost. This makes estimation of the 3D structure of a scene from a set of 2D images very challenging [38]. However, methods such as structure from motion [39], have made 3D reconstruction possible using multiple images with overlapping views. The typical 3D reconstruction process involves inferring the geometrical structure of a scene captured by a collection of images. The camera position and internal parameters are either known or estimated from a set of images. Then the algorithms find the corresponding points in a set of images (i.e., the same point in multiple images) using image features. Finally, using the location of corresponding points in images and their respective camera positions, 3D information can be recovered. Recent development in CNNs has enabled 3D scene reconstruction processes to demonstrate impressive performance in the creation of 3D models from the 2D images [40]. However, 3D reconstruction in construction environments is still a challenging task. This is mainly due to complicated construction environments characterized by poorly textured surfaces that are covered with uniform material, dynamic and complex nature of jobsite, unwanted/obstructed background, repetitive patterns of building surfaces, and occlusion [39].

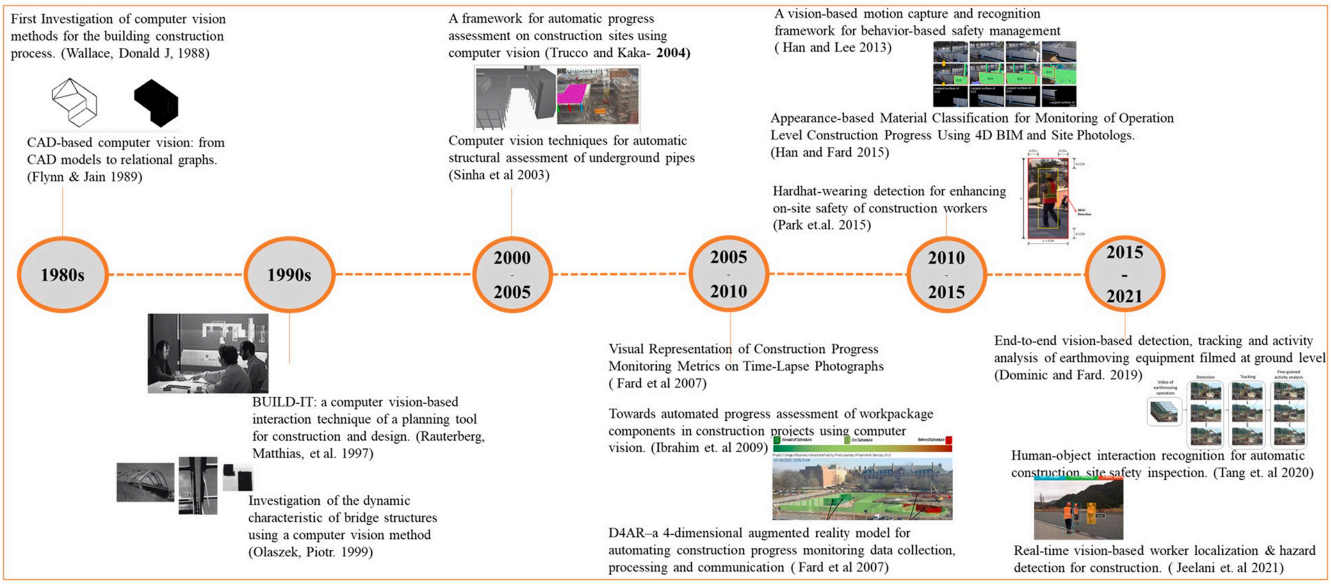

Fig. 2. Evolution of Computer vision in construction: Notable works over the years.

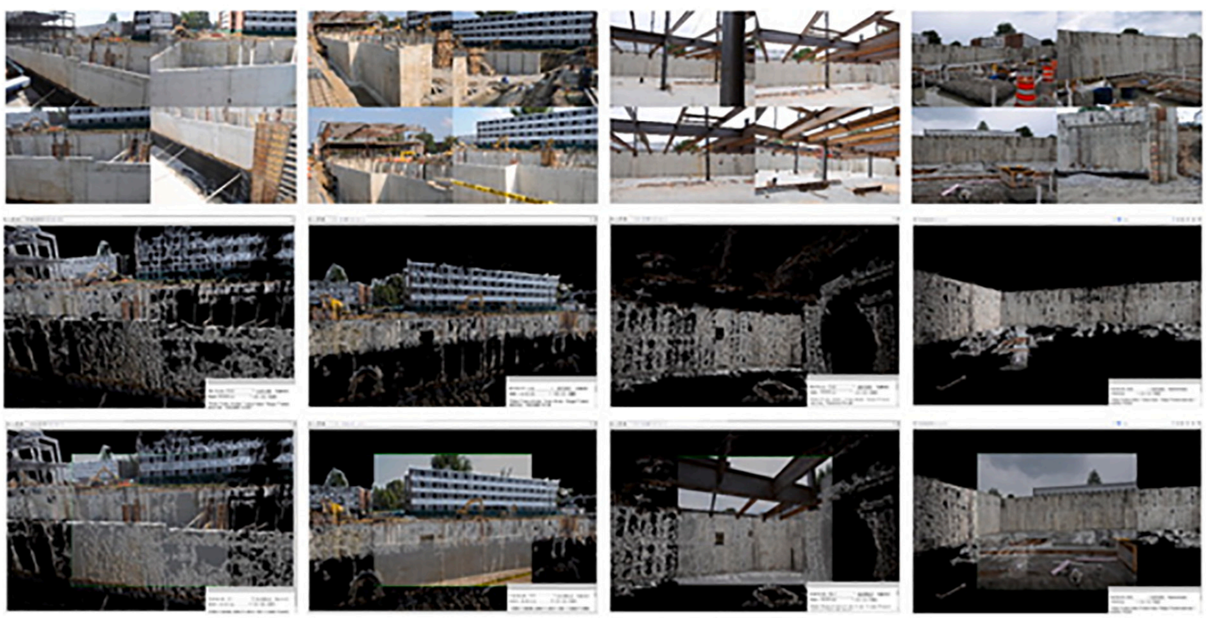

Fig. 3. 3D Reconstruction used for construction progress monitoring by Han & Fard [28].

2.2. Image and object classification

Image classification is one of the fundamental computer-vision tasks of taking in an input image and identifying the '*class*' the image falls under (Or a probability of the image being part of a 'class'). For example, an image classification algorithm can take images of different equipment on a construction site as input and assign a class label such as '*excavator*', '*dump truck*', '*forklift*' etc. to each image (see Fig. 4). Or it can take a single image as an input and provide a probability that the image belongs to a particular class of equipment (such as "*there is a 90% probability that this input is an excavator*"). Sometimes, the image is first broken into discrete objects within them and then each object is classified separately. This type of classification mimics the type of analysis done by humans and is called object classification. It is the process of predicting a specific class to which an object belongs based on object-level features. Object in the context of images is a set of pixels within the image that belongs to the same instance. The object classification technique involves the categorization of pixels based on their spectral characteristics, shape, texture, and spatial relationship with the surrounding pixels. By applying the classification technique, a computer can identify objects into one of the finite sets of classes defined in advance [41]. Classification is one of the core problems in computer vision that, despite its simplicity, has huge practical applications such as the classification of construction workers wearing hardhat, vests etc.

Early image classification methods relied on raw pixel data. The process involved breaking an image into individual pixels and applying statistical methods to categorize images. This, however, is a challenging task as two images of the same class might look very different due to different backgrounds, angles, poses, etc. However, neural networks, particularly Convolutional Neural Networks or CNNs have enabled the development of classification algorithms that identify and extract features from images instead of relying only on pixel data or manual

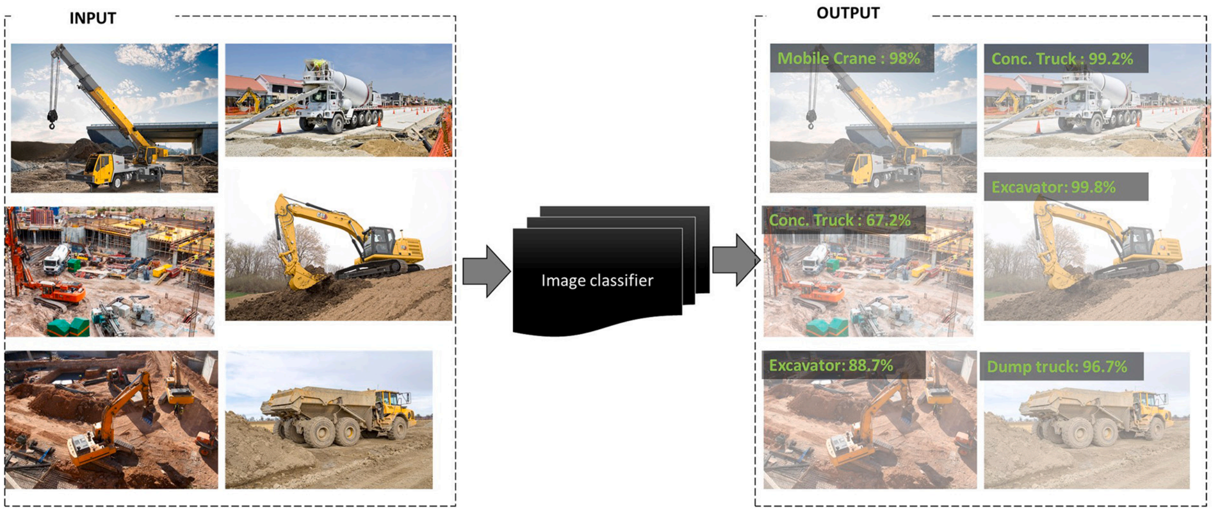

Fig. 4. Classification techniques used for classifying construction equipment and tasks.

application of filters. This has made image /object classification tasks much robust and reliable to be used in a variety of applications. Despite enormous progress, there still exist some challenges to accurate classification due to the object's variability in color, the angle at which the object is located, and speed at which images are taken and most importantly occlusions [42,43].

2.3. Object recognition

As discussed above, Image Classification only provides information about whether (and with what probability) the object of interest is present in an image or not. In contrast, object recognition involves both classification and localization tasks, i.e., it identifies and locates objects in images or videos. It is often a more useful technique as it allows multiple objects to be identified and located within the same image. The object recognition technique recognizes object categories and the location of each object by providing a bounding box encapsulating different objects of interest in an image [41] (see Fig. 5a). An important task in object recognition is to identify what is in the image and with what level of confidence (classification task). Once the object is identified, the next step is to locate it in the image using the detection and segmentation techniques. Detection techniques usually output a rectangle around the recognized object called Bounding Box (BB) (Fig. 5a) generated by regression method [43]. Alternatively, instance segmentation identifies the objects in each pixel resulting in a precise map of the object in an image as shown in Fig. 5b below.

Various machine learning algorithms used for object recognitions are Histogram of Oriented Gradients (HOG) feature extractor [44], Support Vector Machine (SVM) [45], Bag of Features model [46] among others. Advances in convolutional neural networks have made it possible to use object recognition techniques in applications like robotic, navigation, remote sensing autonomous driving, video surveillance, pedestrian detection and several others [47].

2.4. Object tracking

Object tracking is a technique used to track objects as they move across a series of video frames while maintaining their identity and trajectory. The target objects are often people, but may also be animals, vehicles, or other objects of interest. The object tracking process starts with identifying objects and assigns them bounding boxes (i.e., *object detection*). Object tracking techniques assign an ID to each identified object in the image, and in subsequent frames tries to carry across this ID and identify the new position of the same object. Detection of moving objects and motion-based tracking are components for various real-world applications, including pedestrian tracking [48], human-computer interaction [49], autonomous vehicles, robotics, motion-based recognition, video indexing, surveillance and security [50]. Compared to static object detection, object tracking has challenges such as

Fig. 5. Object Recognition output examples: a) Bounding Box b) Instance Segmentation.

i) *Re*-identification—connecting an object in one frame to the same object in the subsequent frames.
ii) Appearance and disappearance—objects can move into or out of the frame unpredictably and we need to connect them to objects previously seen in the video.
iii) Occlusion—objects are partially or completely occluded in some frames, as other objects appear in front of them and cover them up.
iv) Identity switches—when two objects cross each other, we need to discern the two objects.
v) Scale change—objects in a video can change scale dramatically, due to the camera zoom.
vi) Illumination—lighting changes in a video can have a significant effect on how objects look, which can make it harder to consistently detect them.

Some popular algorithms for object tracking that uses deep learning methods are SORT [51], GOTURN [52], and MDNet [53].

2.5. Segmentation

Segmentation is a process of recognizing and understanding what is in the image at the pixel level. The goal of the segmentation task is to give each pixel a label based on what the pixel represents in an image. Thus, images are divided into different regions based on the characteristics of pixels that identify objects or boundaries to simplify an image and analyze it more efficiently. This process allows separating objects from the background. Segmentation tasks can further be categorized as a) Semantic segmentation b) Instance Segmentation.

2.5.1. Semantic segmentation

Semantic segmentation refers to the process of linking each pixel in an image to a class label. Semantic segmentation does not differentiate instances and only uses pixels while providing a richer understanding of an image [41]. Semantic segmentation can be considered as image classification at a pixel level. For example, in the image shown below that, segmentation labels all pixels covering reinforcement bars as green, the ground as purple and structural steel beams as blue without differentiating the individual instances. These computer vision techniques utilize other techniques like object classification (Object detection and localization) to label pixels [54]. Fig. 6 shows an example of scene segmentation.

2.5.2. Instance segmentation

Instance segmentation refers to the process of labelling pixels in an image to the separate instances where an object appears in an image (see Fig. 7). This technique first includes object detection to extract bounding boxes around each object instance, followed by segmentation inside each bounding box to assigns a label to every pixel that corresponds to each instance. Instance segmentation is in a way a combination of object detection and segmentation [41]. The main purpose of instance segmentation is to distinctly represent each instance of the objects of the same class.

2.6. Action recognition

Action or activity recognition is another important computer vision task that aims to recognize the actions of one or more agents from a series of images/videos. Action recognition involves feature extractions from consecutive frames of a video, to identify and classify an action based on a set of predefined action classes, and action localization. The majority of existing action recognition frameworks consist of feature extraction, dictionary learning based on the extracted feature, and classification of video using representation [55]. Since this task requires analysis of a continuous stream of related images (or video), Recurrent Neural Networks (RNNs) are extremely useful in action recognition problems. Mainly there are three types of action recognition techniques as follows [56]. 1) Depth-based action recognition – popular due to the availability of cost-effective sensors. Most existing depth-based action recognition methods use global features such as space-time volume and silhouette information. 2) Skeleton-based action recognition – it uses positions and motion using the coordinates of the joints. 3) Action recognition via a combination of skeleton and depth features – It combines the depth of skeleton and depth features together and helps to overcome situations when there are interactions between human subjects and other objects or when the actions have very similar motion trajectories.

Action recognition techniques are useful in various real-world applications such as security, sports, construction, wildlife etc. In construction, it can be used to determine the actions of various equipment to compute productivity, actions of workers to ensure proper work posture, movement of vehicles and equipment for logistic planning and management etc.

3. Review framework

The review process consisted of the following three steps.

3.1. Identification of literary sources

The first step in the review process was to identify relevant academic journals and databases that publish the latest developments in the field of computer vision in constriction. Journals were selected based on the various parameters such as impact factor, cite scope, Scientific Journal

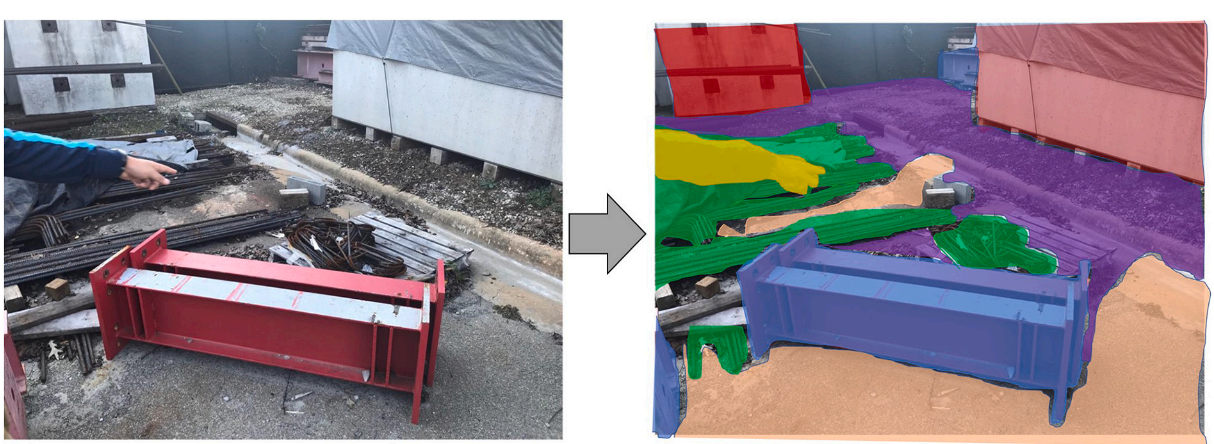

Fig. 6. Semantic scene segmentation.

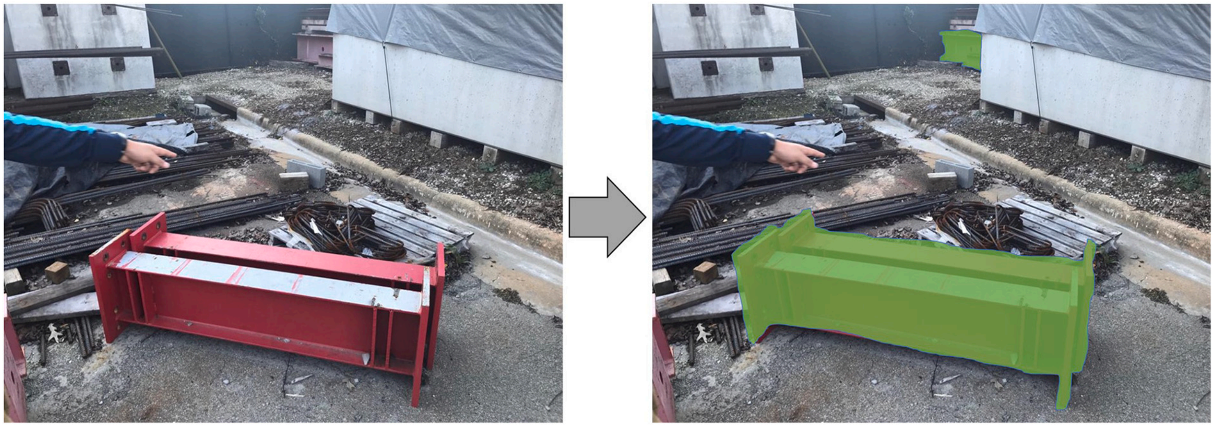

Fig. 7. Instance segmentation (Green masks represent individual instances of an object). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Rankings (SJR) and popularity in the field of computer vision and construction. Databases such as Web of Science, ASCE, and Elsevier were used for the searching journals needed for the review. Table 1 below shows the distribution of articles used in this review.

3.2. Relevant article search

Once the top journals were selected, the relevant articles were searched using different keywords. Google Scholar, a powerful web search engine, was used as a complementary search tool to eliminate searching biases. To include a wide range of construction tasks including safety, quality, progress, and productivity monitoring, the keyword '*computer vision* + *construction*' was used for the primary search that identified most articles reviewed in this study. Further, specific keyword searches such as "*computer vision* + *safety*", "*computer vision* + *progress*", and "*computer vision* + *quality*" were also used to obtain relevant articles in each construction task. More than 550 articles were examined, and the relevant articles were selected using the following inclusion and exclusion criteria for a thorough review.

Inclusion Criteria:

- Work in the construction domain.
- Use of images/ videos as the primary source of data.
- Develop/ use one or more computer vision techniques for a construction-related task.

Table 1
Journal title and articles.

Publication	Journal title	Number of articles reviewed
Elsevier	Automation in Construction	34
	Advanced Engineering Informatics	21
	Engineering	1
ASCE	Journal of Computing in Civil Engineering	20
	Journal of Construction Engineering and Management	13
Hinwadi	Journal of Construction Engineering	1
MDPI	Remote Sensing	2
Sage	Journal of Transportation Research	1
Wiley online	Computer Aided Civil and Infrastructure Engineering	2
Emerald insight	Journal of Construction Engineering	2
Springer	Robotic Fabrication in Architecture	1
	International Symposium on Automation and Robotic in Construction	3
	Conference on Computer Vision workshops	1
	Archives of Computational Methods in Engineering	1

- Work published in journals grouped in the first quartile of SJR rankings.

Exclusion Criteria

- Manuscripts focusing on areas other than construction.
- Use of computer vision in design, asset management, real-estate sales and other non-construction activities within the construction field.
- Manuscripts published before 2010.
- Manuscripts from non-peer-reviewed conferences
- Industry white paper

The refinement resulted in 85 relevant articles. However, more relevant articles were identified from these papers by using reference chains. This helped identify important articles that were missed during the keyword search process. Finally, 101 articles were reviewed thoroughly as shown in Table 1.

3.3. Review and organization

Refined papers were categorized based on different construction management tasks to conduct a focused systematic review. Fig. 8 below shows the number of articles reviewed for each construction task.

Articles were also categorized based on different computer vision techniques as shown in Fig. 9 below.

4. Review findings

Computer vision has drawn attention in construction because of its applicability in automating different construction tasks, monitoring construction sites and automating safety and quality inspections [1,57]. Compared to other sensing techniques such as RFID, GPS and UWB, computer vision techniques have several operational and technical advantages as they can provide information related to one's position and movement with limited sensory data [58,59]. As shown in Fig. 8, safety management and progress monitoring are two leading areas where different computer vision applications have been developed and used over the past decade. This is followed by productivity monitoring. One of the interesting findings was that even though quality control can greatly benefit from advances in computer vision, it is one of the areas that is lagging in the application of computer vision systems. Among different techniques, object recognition seems to be one of the most popular computer vision techniques in construction, closely followed by segmentation and object classification (Fig. 9). The major areas where computer vision has been used significantly and/or has the potential of being used in future:

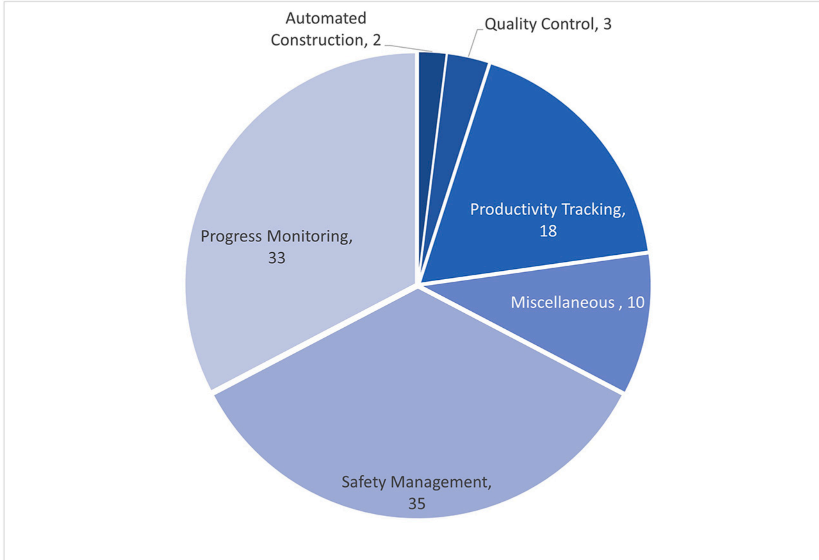

Fig. 8. Articles reviewed- categorized by application area.

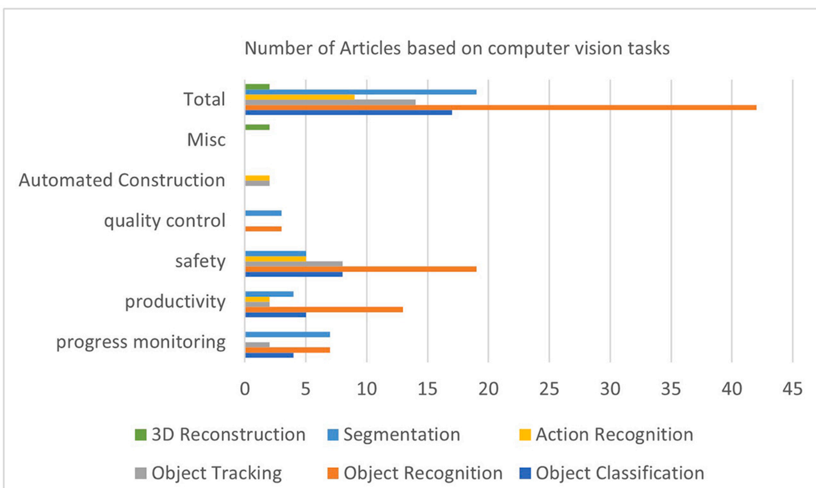

Fig. 9. Articles reviewed- categorized by computer vision tasks.

4.1. *Safety monitoring*

With more than 900 fatal and over 200,000 non-fatal injuries [60], construction is one of the most dangerous industries to work for [61]. Globally more than 60,000 lives are lost on the construction site every year [62]. In addition to the loss of life, and the personal and social impact of injuries, the financial burden of accidents is also significant [63]. To prevent such accidents, it is important to detect unsafe actions and unsafe conditions and take prompt corrective action. The current practice of safety inspection is largely based on human involvement and manual inspections. As with any visual inspection task, computer vision has a huge potential to automate the tasks of detecting and recognizing unsafe actions and unsafe conditions, which is the first step towards eliminating them from the causal chain of accidents.

In recent years, there has been a growing trend in the use of digital tools like Unmanned Aerial Vehicles, predictive analytics, wearable devices in construction sites, which help site safety managers to detect and manage the safety risks that arise during the execution of the project. Computer vision plays an important role in the development and use of these technologies. It can help in identifying and accessing the risk of fatalities and accidents using visual data obtained from the job-sites [26,27,29,64]. Various computer-vision based techniques are used in safety management either individually or in combination to

Table 2
Various computer vision tasks used in safety management.

Method	Literature	Objectives
Object classification	[65–72]	Classifying workers, Risk zones, equipment, and other site hazards
Object recognition	[24,25,27,29,67,68,70,73–82]	Detection of PPE, guardrail, structural supports, equipment, and cranes
Object tracking	[70,77,78,80,83–85]	Tracking workers, equipment, and target
Action recognition	[64,71,85–87]	Recognizing worker's posture, and ineffective human pose; Detecting motion of construction workers
Segmentation	[13,27,66,79,84,88]	Segmenting visual features to detect and/or evaluate safety hazards or safety controls

automatically detect and manage safety hazards on the jobsite. Table 2 shows various computer vision tasks that are predominantly used in safety monitoring, their objectives and a few examples of articles discussing their development and/or use.

4.1.1. Object classification

It is a computer vision technique that has been used to classify objects into predefined groups such as safety conformance and non-conformance, a worker wearing a hardhat and workers not wearing one, vest and no vest etc. Further, the object classification technique has been used to classify construction workers, risk zones, equipment [69], safety behaviors [65], and safety hazards [68] from images collected by surveillance cameras [5], drones [69], ground robots [89] and tablets [24] on construction site. CNNs are often used for such classification tasks. For example, CNN was used to classify workers wearing a hardhat and non-hardhat-wearing workers [72] using HOG features of the hardhats, which are capable of describing detailed shape information efficiently. Similarly, YOLO-V3 [71], which is a single convolutional network that simultaneously predicts multiple bounding boxes and class probabilities for those boxes, was used to measure proximities among construction entities to measure the actual distance of workers from the potential hazards. In addition, workers and non-workers are classified by analyzing spatiotemporal relevance between workers and non-workers. This was done by analyzing, comparing, and matching multiple images of each worker obtained from videos [67]. Finally, color based pixel method has also been used to classify safety vests with a comparative analysis of two colors paces (Lab and HSV [hue, saturation, and value])and three types of classifiers (a support vector machine [SVM], an artificial neural network [ANN], and a logistic regression [LR]) [25].

4.1.2. Object recognition

Object recognition has been used in safety management to detect safety hazards or instances of non-compliance on the jobsite. This includes detection of PPEs to ensure workers are using appropriate PPE [12], detection of guardrails [75] and structural supports [73] to ensure sufficient safety controls are in place detection of static and dynamic hazards such as equipment, vehicles, cranes etc. [24] to aid in safety management at construction job sites. Since it is not possible to eliminate all hazards at a job site, PPE compliance becomes very crucial. PPE protects workers from exposure to potential hazards that cannot be eliminated by engineering and administrative controls. Various object detection and recognition techniques have been used to help identify whether proper PPEs are being used by workers. For example, [25] used a background subtraction method and color pixel classification to detect safety vests on workers, which was subsequently used to identify workers not wearing an appropriate safety vest. Similarly, [12,74] developed a vision-based system to detect hard-hats on workers using pixel-based image classification. In addition, vision-based methods using CNNs have also been used to detect workers to obtain the positional information of workers, and with the help of positional information, danger zones for the workers were detected [65]. Falls contribute to about one-third of all fatalities in construction and the use of a safety harness and appropriate PPE is critical to reducing the risk of falls [90]. Object recognition techniques can greatly help in monitoring compliance to appropriate fall protection systems. For example, [81] used Faster R-CNN to first detect the presence of workers on heights and then CNN based classification algorithm to determine whether or not workers are using the safety harness. Similarly, [73] used CNN based occlusion mitigation method to monitor PPE compliance by steeplejack workers working in the exterior wall. In addition to PPEs, barriers such as guardrails can minimize the workers' exposure to hazards. Computer vision can play an important role in ensuring that necessary guardrails are in place. For example, [75] developed a method to detect guardrails using Visual Geometry Group architecture (VGG-16). Further, R- CNN was used to detect structural supports present on construction sites [24,27]. This subsequently helped to reduce the fall hazards. Furthermore, Improved Faster Regions with Convolutional Neural Network (IFaster R-CNN) approach was used to automatically detect the presence of excavators and workers on the site with a high level of accuracy (91% and 95%) [24]. These detections help in safety monitoring by automatically detecting the worker's unsafe proximity to a hazard. Finally, each construction trade has its own set of safety standards and codes to follow. So, identification of trades also becomes crucial for efficient safety management. Moreover, sometimes it becomes necessary to designate specific zones to trades to minimize exposure to hazards. Computer-vision can greatly help in ensuring compliance to such measures. For example worker's certification checking system developed via video imaging based on R-CNN was used to identify types of trade working on the site [67].

4.1.3. Object tracking

Object tracking is another computer-vision technique that is very helpful in tracking workers [77], equipment [78], site dynamics [85], and motion characteristics of onsite objects [70] to ensure sufficient safety measures are in place and workers are maintaining a safe distance from the potential hazards. Various algorithms such as mean shift, Bayesian segmentation, active contour, and graphs algorithm are used to track workers using moving cameras. For example, the Bayesian method, which uses the segmentation procedure, was used to track workers on the construction site [83]. This approach offers advantages over gradient-based methods as the segmentation process is global and pixelwise. Workers are also detected and tracked using pedestrian detection techniques [77] from jobsite's bird's eye view images. First, a decision forest algorithm is used to detect workers, then the soft cascading classifier is used to track workers [77]. Similarly, workers can be tracked from a video of the jobsite using a combination of detection and tracking algorithms. For example, [78] combined latent SVM detection algorithm to first detect workers and then used particle filtering to track the detected workers. Also, [80] used HOG template algorithm to detect workers then used template matching to track workers by using the region from the detection bounding box. Tracking equipment and its movement in the jobsite remain a challenging task on construction. A 3D spatial modeling algorithm and image matching algorithm was used to track equipment and its surrounding [70]. This helps in safe equipment operation by providing accurate information about objects surrounding the equipment. Since jobsite surveillance cameras are already in use on the most construction sites, a video-based tracking system can be implemented at low costs using these cameras.

4.1.4. Action recognition

Action recognition has been used in construction safety to identify an action or task that workers are engaged in, by recognizing workers motion, position, and body movement. Multiple computer vision algorithms are used to identify these actions. For example, to determine a worker's pose used a Linear discriminant analysis (LDA) classifier to classify workers pose from the images collected by a range camera [87]. This helped determine whether a worker is standing, bending, sitting, or crawling. Then OpenNI [91] middleware was used to get the joint angle and spatial locations. Finally, body posture information was used to categorize tasks as ergonomic or non-ergonomic [87]. In another study, motion capture data from Kinect was used to identify the unsafe actions of workers climbing a ladder (i.e. backward facing climbing, climbing with an object, and reaching far to a side) with an accuracy of over 90% [92]. In addition, 3D skeleton extraction and motion recognition technique are used to detect unsafe actions such as reaching too far in ladder climbing [64]. Images taken from smartphones can also be used to capture human motion data for onsite motion sensing and analysis [85], which ultimately helps to detect the action performed by the worker. Worker's motion and position can also be analyzed from the video obtained from the jobsite to identify non-ergonomic postures and movements by acquiring 2D skeleton and 3D coordinates of joints by extracting image sequences [86].

4.1.5. Segmentation

Segmentation has been used in safety management to locate visual features from images including color, texture, compactness, contrast, and edges, which ultimately enables the identification of safety hazards. For example, the background subtraction algorithm extracts motion pixels from an image sequence or video, pixels are then grouped into regions to represent moving objects, and finally, the regions are segmented using classifiers [66]. The connected moving regions are used for object correspondence and classification. This technique provides the benefit of site monitoring with reduced human intervention by using the video obtained from the site. In another study, the unsafe behavior of workers around structural supports in a deep-pit foundation was detected and segmented using Mask R-CNN [27]. The overlapping detection model was trained to determine the relative position of workers and structural supports [27]. Additionally, [88] used instance segmentation using MaskRCNN to develop a vision-based system that provides real-time alerts to workers if they are in proximity of hazards. Segmentation technique has also been used to detect hard hats of workers to check compliance [13]. In addition segmentation, feature representation, and classification technique can also be used to detect high-risk areas of a construction project such as roofs, edges etc. [79].

4.2. Progress monitoring

Progress monitoring is one of the most important tasks in construction management as it tracks the progress of the project and ensures that the project is constructed on schedule and within budget. An accurate assessment of progress allows managers to make better decisions to control the project's cost and schedule. Current practices of progress monitoring require significant manual intervention, are time-consuming and are prone to human errors. Therefore, efficient monitoring systems can help construction teams by automating progress inspections, which will help to reduce the risks of reworks and errors and prevent deviations of cost and schedule [14]. To reduce reworks and errors, construction companies often integrate their production schedule with the 3D BIM to create 4D BIM. However, this process involves manual manipulation to integrate real-time progress information with 4D BIM. To facilitate this integration process, construction researchers and practitioners have focused on collecting as-built visual data through hand-held cameras and video recorders, assigning field engineers to filter, annotate, organize, and present the collected data in comparison to as planned date from 4D BIM. However, the cost and complexity associated with manually collecting, analyzing and reporting operations results in sparse and infrequent monitoring and a portion of the gains in efficiency are consumed by monitoring costs [6]. Therefore, construction researchers are working towards automating such manual processes. Advances in the field of computer vision have enabled the development of technology that has aided in automating various tasks involved in progress monitoring.

Typically, progress monitoring using visual data requires capturing as-built data, which can be in the form of still images, videos or point clouds, and comparing this with the as-planned models from BIM, CAD etc. This work predominantly requires one or more of the following computer vision techniques as shown in Table 3

4.2.1. Object recognition and classification

This technique has been used to detect building elements to compare planned vs actual finished work [3] by analyzing the visual data obtained from the site via drones or other cameras to evaluate the progress of the project. For example, work packages were generated automatically by analyzing multiple images obtained from the site, which allowed project controls to be monitored effectively [93]. In another work, site photographs were compared with the BIM model to generate the status of interior construction by decomposing the as-built model [95]. Object detection was then used to detect interior construction elements and subsequently matched with interior construction images to identity progress schedule deviation. Similarly, in [98] the progress of interior construction was evaluated by using integrated shape and color-based modules, which detected studs, insulation, electrical outlets, and different states for drywall sheets (installed, plastered, and painted). Based on the results of the modules, images are classified into one of five states of construction i) framing, ii) insulation, iii) insulated drywall, iv) plastered drywall, and v) painted partition [98]. Moreover, object detection has been used to detect building elements and obstacles, and visual comparison with BIM model to assess progress. For example, appearance and geometrical based reasoning was used to evaluate construction progress using Earned Value Analysis (EVA) concepts from an integrated model of point cloud and BIM [6]. Similarly, a probability distribution algorithm was used to recognize material appearance by the filter bank and principal Hue- Saturation color vale with an accuracy of 97% [7]. This helped to identify changes on jobsite based on the material appearance. In addition, 3D structural components can also be recognized using color data and stereo vision systems to automatically track structural progress in construction [97]. Furthermore, with the help of geometrical model matching and statistical analysis of template mask regions, progress can be monitored using image-based classifiers, which identify changes and detects construction processes [100]. In another work, [99] developed texture-based reasoning for image-based 3D point clouds and color-based reasoning for laser-scanned point clouds [99]. In this work, geometry-based filtering detects the state of construction of BIM elements (e.g., in-progress, completed) and appearance-based reasoning captured operation-level activities by recognizing different material types [99].

Table 3
Various computer vision tasks used in progress monitoring.

Method	Literature	Objectives
Object classification	[93–96]	Compares planed vs actual work, tracks status of construction and progress
Object recognition	[6,7,97–100]	Detects building elements and obstacles, visual comparison with BIM model to identify progress
Object tracking	[101,102]	Track objects in real time
Segmentation	[4,6] [8,93,97,98,101]	Segments visual features including color, texture, compactness, contrast, edges

4.2.2. Segmentation

Scene segmentation can help to understand the contents of the image to determine the progress of the project. For example, the voxel coloring algorithm was used to monitor the progress of building elements by comparing 3D obtained from images of the site with the BIM model using the Bayesian machine learning model [8]. This comparison generates construction progress deviation with a color code. The appearance-based method can also be used in image-based as-built models to extract image patches by back projecting RGB images [6]. Similarly, [4] developed an automated material tracking system to track the productivity of the project; whereas [93] used segmentation to automate the tracking of work packages. [98] have also developed techniques for automatically detecting building components such as studs, insulation, electrical outlets, and drywall sheets.

4.2.3. Object tracking

By tracking objects and their positions, the change in the task can be determined, which ultimately helps in evaluating progress. In addition, object tracking can help in autonomous data-collection useful for progress monitoring. For example, localization algorithms such as Simultaneous Localization and Mapping (SLAM) can be used to track and navigate robotic systems that autonomously collect visual data to compare as-built information with the as-planned work [101].

4.3. Productivity tracking

Productivity is defined as a total output per unit input and is usually expressed as the cost of labor or man-hours. Globally, labor productivity growth of construction lags behind that of manufacturing and the total economy [103]. As such, measuring productivity has become an ever more important task in construction. Labor and equipment productivity is a key indicator of project performance [104]. It helps to optimize resource planning, which is critical to counter the challenges of the decreasing supply of labor [103]. In addition, most construction projects have tight budgets with a very thin profit margin of around 5% [105] and, as such, require highly optimized resource utilization making productivity monitoring extremely vital. The monitoring of productivity for different construction processes remains a difficult and error-prone task as significant manual effort is often required to measure the output of labor, equipment or resources [106]. Therefore, automating the process of productivity monitoring can help the project team achieve optimum utilization of resources with minimal human intervention. Vision-based techniques can help automate and track productivity [106] by analyzing the workers' movements and interactions [106], action analysis [107,108], and location tracking [109]. For example, interior construction's productivity was improved by analyzing various quality checks and inspection reports [95,109,110] that track workers movements and interactions on site. Various computer-vision based techniques have been used in productivity monitoring either individually or combined as shown in Table 3 and Table 4

4.3.1. Object classification

Operational level deviation during the construction hinders productivity. To measure productivity at the operational level, construction managers use pictures, videos, and daily logs. Object classification has the potential to be used as an efficient tool to monitor such progress deviations. Image-based point clouds built from images and videos that are already collected from jobsites can be used to extract information about productivity without adding a new burden of requiring expertise for data collection and analysis [15]. For example, 3D point clouds generated from site images using Structure-from-Motion techniques were used to track project deviation by integrating the production schedule [111]. Also, improved CNN was used to monitor worker's activities in concrete construction [33]. This helped to measure man-hours and as well as the productivity of each worker. Progress status of construction can also be evaluated by comparing as-planned BIM models and 3D models obtained from as-built photographs, which helped to track productivity by action analysis [95]. Various algorithms such as Histograms of Oriented Gradients (HOG), Histograms of Optical Flow (HOF), Motion Boundary Histogram (MBH) can also be used for action learning and classification, which can also help in tracking productivity [112].

4.3.2. Object recognition

Object recognition is another technique that has been used to evaluate the performance of action analysis and detection of construction workers and equipment to assess productivity at various levels [104]. For example, high-resolution satellite images were used to extract features to detect various objects of interest. The detected objects, integrated with spatiotemporal database and baseline schedule, automatically provided the location-based progress data [115]. This helped to ensure that specific areas of the project are meeting the desired level of productivity. In addition, Bayesian learning and Bag-of-Video-Feature Words models were used to recognize worker's movement and machine movement to measure individual productivity levels [114]. In another interesting study, HOG was used for the pose estimation of excavators. Then a spatial-temporal reasoning model was used, which uses time and space constraints of the excavators' moving patterns to measure the productivity of excavators from videos [116]. Similarly, support vector machine (SVM) classifiers were used to recognize equipment movement actions using visual features such as space-time interest extracted using HOG algorithm [118]. Some other techniques used to track productivity include automated image-based reconstruction and modeling of the as-built project status using unordered daily construction photo collections through analysis of Structure from Motion (SfM) [109].

4.3.3. Object tracking

Object tracking has been used to track the motion and movement of equipment in construction. By tracking the motion of objects, productivity can be evaluated by analyzing its cycle time and working processes. For example, Histograms of Oriented Gradients and Colors (HOG+C) algorithms [117] were used to detect workers and equipment from the videos obtained from the construction site. Similarly, the video computing method was used to automatically detect and track the project resources, work state classifications, and production scenarios such as working processes, cycle times, and delays with an accuracy that of manual analysis [106,107,120]. The results indicated the promising effectiveness of the automatic video-based method to measure productivity compared with manual processes.

4.3.4. Action recognition

Action recognition is an important computer vision task that helps in tracking productivity by recognizing the action of equipment or workers. Although this largely remains an unexplored area, there have been a few works that have used action recognition for productivity analysis. For example, the probability graph model was used to estimate the jib angle of a crane to analyze if the task is concrete pouring or other material movements [108]. Similarly, Bag-of-Words and Bayesian network models were used to learn and classify actions of construction workers and equipment to subsequently identify their work tasks [114].

4.3.5. Segmentation

The segmentation technique is used to track productivity by analyzing the object's pixel value based on the color thresholds. For example, the background subtraction method was used for segmenting static background on a video sequence of a crew installing formwork and earthwork [119]. The helped to measure the productivity of formwork and earthwork activities. In addition, segmentation was used to detect pile caps by analyzing images obtained from the jobsite [115]. Haar-HOG was used to extract visual features from the time-lapse videos, to measure the productivity of dump trucks in the construction site [110].

Table 4
Various computer vision tasks used in productivity tracking.

Method	Publication	Objectives
Object classification	[15,33,95,111,112]	Monitoring progress deviations at the operational level,
Object recognition	[11,104,106,107,109,110,112–117] [118]	Progress tracking based on the location, Evaluating the performance of action analysis, Detection of construction workers and equipment
Object tracking	[104,108]	Tracking motion and movement of equipment
Action recognition	[104,114]	Detecting workers interactions between actions and related objects
Segmentation	[114], [110,115,119]	Segmenting visual features of a construction element

4.4. Quality control

Construction projects often experience cost and schedule overruns and rework is one of the factors that contribute to these overruns [121]. Reworks are the products of quality deviations, nonconformance, defects, and quality failures, and result in the unnecessary effort of redoing a process or activity that was incorrectly implemented the first time. The direct costs of rework are approximately 5% of the total construction costs [121]. To minimize the reworks, quality control (QC) managers have developed various QC programs; however, they are mostly manual and need a significant involvement of human resources. For example current method of dimensional analysis of construction, components are based on the use of remote-sensing instruments such as Total Stations [122], which needs several manhours for both field and office work to capture field information and to process it. Computer vision can capture not only dimensional information but also spatial information. Integration of BIM and computer vision can thereby aid in QA/QC by evaluating dimensions, plumb, installations etc. However, current practices of quality control using BIM are still labor intensive. Therefore, computer vision techniques need to be explored to automate the processes that can increase the efficiency of quality control and reduce human effort.

This area within construction has largely lagged in the adoption of computer vision techniques. However, there have been some preliminary explorations in this area that have been listed in Table 5 below. Mostly object recognition and segmentation techniques have been used to detect quality defects such as cracks, misalignment, and dimensional discrepancies using visual features, texture, color, and edges. For example, the alignment inspection of tile installation was improved by analyzing the geometric characteristics of the finishes of the tile surface [123]. In addition, geometric and relationship-based reasoning was used to check dimensional discrepancies by comparing with as-built and planned BIM to automatically identify dimensional discrepancies [124]. Segmentation of multi-scale feature detection was also used to detect the surface and curvature of objects [125]. Even though the use of computer -vision has been somewhat limited in quality control, there are numerous opportunities to explore different vision-based techniques to examine evaluate built structures for defects and non-compliance, which need to be explored in future.

5. Challenges & opportunities

In 2012, Forbes reported that jobs in the skilled trades were the most difficult to fill in the United States and as of now in 2021, the problem not only persists but is exasperated by a historically low economy-wide and construction-specific unemployment rate [126]. Therefore, there is an ever more critical need for automation in construction. Automated construction can help to fill the skill gap that exists in construction skilled trades. Although there has been slow progress in the development of technology required for automated construction, there is likely going to be a spike in research in this area in near future. Already automated robots have been used for the task of painting and bricklaying [127]. Several other tasks that are manually performed in the construction site can be automated and computer vision will have a huge role to play in this transition. Specifically, the sensing/actuation feedback loop can be used to predict the construction environment where machines, materials and human being interact. For example, with the help of Scorpin, a robot control plugin, multiple layers of sensing provided a feedback loop in a bricklaying robot [128]. In addition, a vision-based algorithm was used to assist the robotic system for the quality monitoring system [109]. The system can automatically adjust the extrusion rate based on the feedback from the algorithm and would be able to print layers of acceptable dimensions using a printable mixture, without the need for prior calibration and despite mixture rheology variations. The high precision and responsiveness of the developed system demonstrate the great potential for computer vision as a real-time quality monitoring and control tool for robotic construction [23].

In addition to autonomous or robotic construction, enormous opportunities exist for computer vision to be explored for safety, progress, quality, and productivity monitoring. These opportunities include developing new and innovative platforms using various vision-based techniques to provide real-time and high accuracy tools that can automatically create information-rich digital twins of construction sites to provide useful insights about progress, quality, and safety. Most of the current studies that have used computer vision in these areas are exploratory and are not currently at the level where these systems can be deployed on real-construction sites efficiently and economically. Future research efforts need to build on these exploratory studies and invent ways to scale up these systems for practical implementation. For example, in safety management, research needs to focus on building computer-vision based techniques to efficiently locate workers and equipment in real-time using live images/videos to provide real-time proximity warnings. There are also opportunities to develop customized network architecture for construction use and build a comprehensive visual database of hazards that can be used to train neural networks to detect different types of safety hazards on job sites. In progress monitoring, research needs to focus on building image processing or AI-based algorithms to efficiently compare as-built images with as-planned models. There is an immediate need to research an efficient way to align (in terms of scale, rotation, and translation) an image-based 3D model with as-planned BIM model. For productivity analysis, the latest developments in recurrent Convolutional neural networks (RCNNs) must be exploited for action recognition and tracking of workers and equipment from live videos. This can help in developing an efficient system for autonomous progress monitoring. Among others, quality control offers the greatest number of opportunities for computer vision-based systems and ironically this is the area where the least studies have been published. The use of computer vision for quality control must be explored as it offers numerous advantages over traditional methods. It saves human effort (and subsequently cost) and prevents errors arising from human factors such as inattentiveness, exhaustion, or simple boredom. Furthermore, to develop techniques and algorithms for defect detection, future research should also adopt techniques and methods from the manufacturing industry and focus on developing vision-based quality control techniques that use live images collected during construction (instead of post-construction quality checks). This will help in minimizing the defects and reduce the need for rework.

Although research has demonstrated huge potential for computer vision in various areas of construction and project management, this review also highlighted some challenges. Perhaps one of the biggest challenges in implementing computer vision is the lack of a visual dataset specific to construction environments needed to train different neural networks [129–131]. There have been multiple efforts to develop several benchmarking datasets such as ImageNet, KITTI, MOT, Cityscapes etc. However, collecting a large amount of annotated data is not an easy endeavour. In addition, there are other challenges specific to the domains and the type of techniques used. Table 6 summarizes the key challenges of computer vision in construction progress monitoring, safety, and quality control. These challenges are explained further to illustrate how they affect the implementation of computer vision in construction. For example, [25] reported some of the important challenges in implementing computer vision for action recognition of construction workers or equipment. These include (1) lack of datasets (2)

Table 5
Computer vision techniques used in quality monitoring in construction.

Method	Publication	Objectives
Object recognition	[123–125]	installation verification, as-built schedule, and dimensional discrepancies
Segmentation	[123–125]	Segments visual features including color, texture, compactness, contrast, edges

Table 6
Challenges of computer vision in construction.

Construction task	Challenges	Publication
Progress monitoring	Lack of detailed BIM model	[6,132]
	Automatic integration with building information model (BIM)	[96,98,102]
	Occlusion & limited visibility	[4,6,15]
	Lack of formalized construction sequences	[132]
	Integrated cost-schedule control systems	[133]
Safety	Lack of annotated dataset	[6,27,29,134]
	Obstacles in the construction jobsite	[12,24,134]
	Occlusion & Limited Visibility	[134]
	Privacy concern to monitor the construction site	[134]
	Workers with different body postures	[134]
Quality control	Surface roughness	[125]
	Curvature of surface	[125]

complex actions of construction equipment and workers; (3) lack of knowledge to define a time-series of actions; (4) simultaneous action recognition of multiple project entities; and (5) lack of a holistic approach to benchmarking, monitoring, and visualization of performance information. Similarly, for progress monitoring, one of the most common challenges in implementing computer vision-based systems for progress monitoring reported in the literature is the missing information on the planned 3D model.

5.1. Data challenges

The lack of annotated dataset is one of the biggest challenges to implement deep-learning-based computer vision techniques in construction. Even though there are several publicly available datasets such as ImageNet [135] and Microsoft® Common Objects in Context (COCO) [136], datasets required for construction processes need to consider unique characteristics such as cluttered backgrounds, occlusions, various poses and scales and the dynamic nature of construction environment. Construction jobsites are often complex and dynamic, with every site uniquely different. This demands a comprehensive dataset that can address complexity, dynamics, and variations in our industry. Although different datasets have been created and used for various applications, these have often been created individually, on a small scale with limited sharing capabilities. Efforts need to be made to create a combined and large comprehensive labelled visual dataset that can be used by construction researchers and developers for diverse applications.

In addition to the lack of a visual dataset, implementation of computer vision in construction has also been hindered due to the unavailability or inorganization of other data needed for semantic understanding of visual data. These include:

5.1.1. Collecting good quality data

Computer-vision based applications rely on analyzing the collected visual data (i.e., images and videos). Therefore, the quality of that input data impacts the performance of the system the most. For example, an efficiently designed object detection architecture, pre-trained on good quality data can severely underperform if the input images are of poor quality (e.g., blurry or collected from awkward angles). Several factors can affect the quality of input data. While some can be controlled such as camera position, orientation, and stability; others are sometimes beyond our control such as poor lighting, cluttered backgrounds, occlusions etc. The upfront cost of the hardware infrastructure and the recurring costs of personnel required to ensure a good quality data collection is another challenge that hinders the implementation of computer-vision based systems for construction applications.

5.1.2. Lack of detailed BIM model

For automatic registration and accurate comparison of as-built images (or point clouds), planned BIM model needs to have significant details and should at least be at the Level of Development (LOD) 400/500, which reflects operational details within the work break down structure (WBS) of the schedule [132]. Current BIM models often lack these important details. For example, scaffoldings or formworks do not feature in as-planned models, therefore comparison between as-built visual data (that contain these elements) with as-planned becomes challenging. Sometimes re-projecting the position of these elements using different machine learning techniques can overcome this challenge to some extent [94]. In addition, daily operation-level tasks are also not reflected in the work breakdown structure (WBS). This lack of a detailed model prohibits accurate comparison between as-built and as-planned models using vision techniques. For example, when planned BIM model of basic geometrical details (LOD 200) is compared with the as-built model obtained from visual data from the site (LOD equivalent of 500), progress metrics (such as per cent complete) obtained using vision techniques may not accurately reflect the actual progress of work.

5.1.3. Automatic integration with building information model (BIM)

A progress tracking system uses the current work as a means of detecting the actual state and compares it with the 4D BIM. In other words, progress is detected by comparing as-built 3D models with as-planned BIM models. This requires the models to align perfectly (aka registration). Even slight inaccuracies in the registration process can compromise the accuracy of calculated progress. As we do not have a robust method to register models automatically and accurately, this step is commonly performed manually. This increasing the human intervention, thereby increasing the time, cost and error proneness in using CV for progress monitoring [102]. There have been some efforts to automate the process [124] but more work is required in this area. For example, the automatic registration of 3D data from BIM to a 3D model of the structure (as-built) the Levenberg-Marquardt algorithm has been used to automate the process with some success [96].

5.1.4. Lack of task-specific quantitative metrics to evaluate unsafe conditions and acts

Construction projects use diverse metrics for quality and safety assessment which often vary between the projects. In addition, most of these metrics are qualitative and it is challenging to quantify these qualitative metrics [134], which is necessary to train a vision-based algorithm. This lack of well-defined and consistent metrics hinders the implementation of computer vision techniques for quality and safety monitoring.

5.1.5. Lack of formalized construction sequences

In several cases, construction projects lack formalized construction sequence that defines all the operational level details. Often only high-level activities are defined in the schedule, which limits the effectiveness of a vision-based technique.

5.2. Occlusion and limited visibility

Occlusion is the effect of one object in a 3-D space blocking another object in view due to the visual and lighting conditions of the environment [39]. In construction, the visual environment is often cluttered with varied lighting conditions that exacerbate the problem of occlusion. The complexity of construction site, obstruction, lighting level, equipment size, color and shape of worker's clothing [12] add to the challenges of occlusion and limited visibility. In addition, construction sites also consist of objects involving shiny, reflective, or glossy surfaces, which also pose a challenge. There have been some research efforts to overcome these challenges. For example, [137] used Fermat paths of light between a known visible scene and an unknown object not in the line of sight of a transient camera to create a prediction of hidden surfaces. However, occlusion and limited visibility issues have not been resolved completely and continue to be a challenging obstruction in implementing computer vision in construction. Efforts such as capturing images from different viewpoints – assisted by the UAVs– can overcome some of the challenges associated with limited visibility of construction elements, yet occlusion cannot be eliminated. Therefore, there is a need to account for occlusion and limited visibility in the design and development of vision-based systems [6].

5.3. Privacy issue in Jobsites

Visual data in construction can be obtained via drones, ground robots, and mobile devices. However, the privacy of data remains a concern for workers whose activities get captured by these visual sensors. The question of who owns data and how the data can be used remains unclear. Moreover, although the objective of visual data collection is not worker monitoring, their activities inevitably get recorded in the process of data collection. This may result in workers feeling under constant observation, causing levels of anxiety and stress that could have an adverse impact on their mental health. Future research efforts need to explore data privacy and worker health and safety issues related to the collection and use of visual data for different computer vision techniques.

5.4. Variations

In addition to the above challenges, variations in the appearance of objects or actions of interest in construction present an important challenge that limits the effectiveness of computer vision techniques. These variations include intra-class variation (E.g., same equipment from different brands can appear very different), scale variation (same construction elements such as walls can be of varying sizes), view-point variation (elements might look different depending on where the camera is located). These variations often result in misclassification. For example, a vision-based system designed to detect worker posture might misclassify the results as workers have different ergonomics and as a result, their body posture while crouching down, bending, and sitting, differs for individual workers. In addition, construction surfaces can vary as well. The surfaces can be flat, single curved, double curved, or have undulations at multiple scales, making boundaries hard to define. These variations in surface geometry also add to the complexity involved in different computer vision tasks. In addition to the geometry, the physical texture of common surfaces can range from smooth (steel, marble) to very irregular (grass, crushed stone), which also presents a challenge for computer-vision based techniques.

5.5. Semantic gap

Most of the applied computer-vision techniques learn from correlations and/or recurring patterns in the input data (specifically between the features extracted from images). Unlike humans, computer-vision algorithms can seldom draw causality or extract higher-level semantic understanding from images (or videos). While this semantic gap may not be an issue for some applications (e.g., detection of cracks in a column), it limits the application of automated computer-vision based systems in areas where context is important [31]. For example, safety management applications require a computer vision-based system to not only detect objects but also evaluate the interaction between objects (e.g., worker and hazards). This high-level semantic understanding is often challenging as it requires significant domain knowledge to be encompassed in the system in addition to training it for detection and analysis of low-level image features.

6. Conclusion

The construction industry captures tremendous amounts of visual data daily to track and control project progress, safety, and productivity. However, extracting meaningful information from pictures and videos necessary to make decisions, is a difficult task and requires a high level of human involvement. Recent advances in computer vision have enabled the development of techniques and systems to automate information extraction and augment data-driven decisions. This review offered an updated and categorized overview of computer vision applications in construction from a holistic approach and identified opportunities and challenges that future research needs to address to fully leverage the potential benefits of Computer Vision in construction. The review indicated that safety management and progress monitoring are the most popular fields that use computer vision followed by productivity tracking and quality control. Among different computer vision techniques, object recognition is the most used technique that has found several applications in construction management followed by segmentation and classification. Object tracking and action analysis have been quite popular in safety management but have not been explored much in other fields despite promising applications, especially in productivity monitoring.

The review also resulted in the identification of various challenges that hinder the implementation of computer vision-based systems in construction. One of the most common challenges is the need for good quality input data for vision-based applications to work efficiently. Owing to the unique characteristics of construction environments (uneven ground, poor lighting, cluttered environments), maintaining this good quality of visual data is difficult. The second most common challenge reported by several researchers is the lack of adequately sized labelled or annotated data needed to train AI-based systems. The literature suggests that there is a lack of extensive databases of labelled images that can be used to train vision-based systems for construction applications. In addition, lack of detailed BIM models, complex, dynamic and occlusion-prone environments and variations in appearance of objects (material, equipment, building elements), variation in task sequence and methods are some of the other challenges that are inhibiting the development of computer vision in construction. These challenges need to be addressed or accounted for when developing vision-based systems for construction applications.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. OAC-2004544.

References

[1] Kevin K. Han, Mani Golparvar-Fard, Potential of big visual data and building information modeling for construction performance analytics: an exploratory study, Autom. Constr. 73 (2017) 184–198, https://doi.org/10.1016/j.autcon.2016.11.004.

[2] B. Akinci, F. Boukamp, C. Gordon, D. Huber, C. Lyons, K. Park, A formalism for utilization of sensor systems and integrated project models for active construction quality control, Autom. Constr. 15 (2) (2006) 124–138, https://doi.org/10.1016/j.autcon.2005.01.008.

[3] Jun Yang, Man-Woo Park, Patricio A. Vela, Mani Golparvar-Fard, Construction performance monitoring via still images, time-lapse photos, and video streams: now, tomorrow, and the future, Adv. Eng. Inform. 29 (2) (2015) 211–224, https://doi.org/10.1016/j.aei.2015.01.011.

[4] Danijel Rebolj, Nenad Čuš Babič, Aleš Magdič, Peter Podbreznik, Mirko Pšunder, Automated construction activity monitoring system, Adv. Eng. Inform. 22 (4) (2008) 493–503, https://doi.org/10.1016/j.aei.2008.06.002.

[5] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T.M. Rose, W. An, Detecting non-hardhat-use by a deep learning method from far-field surveillance videos, Autom. Constr. 85 (2018) 1–9, https://doi.org/10.1016/j.autcon.2017.09.018.

[6] K.K. Han, J.J. Lin, M. Golparvar-Fard, "A Formalism for Utilization of Autonomous Vision- Based Systems and Integrated Project Models for Construction Progress Monitoring," Proceedings of the 2015 Conference on

Autonomous and Robotic Construction of Infrastructure, Ames, Iowa. https://lib.dr.iastate.edu/intrans_reports/141/, 2015.

[7] Andrey Dimitrov, Mani Golparvar-Fard, Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections, Adv. Eng. Inform. 28 (1) (2014) 37–49, https://doi.org/10.1016/j.aei.2013.11.002.

[8] M. Golparvar-Fard, F. Peña-Mora, S. Savarese, "Monitoring Changes of 3D Building Elements from Unordered Photo Collections," IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 249–256, https://doi.org/10.1109/ICCVW.2011.6130250.

[9] Julie M. Sampson, Sarah DeArmond, Peter Y. Chen, Role of safety stressors and social support on safety performance, Saf. Sci. 64 (2014) 137–145, https://doi.org/10.1016/j.ssci.2013.11.025.

[10] Christian Koch, Kristina Georgieva, Varun Kasireddy, Burcu Akinci, Paul Fieguth, A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure, Adv. Eng. Inform. 29 (2) (2015) 196–210, https://doi.org/10.1016/j.aei.2015.01.008.

[11] F. Spencer Billie Jr., Vedhus Hoskere, Yasutaka Narazaki, Advances in computer vision-based civil infrastructure inspection and monitoring, Engineering 5 (2) (2019) 199–222, https://doi.org/10.1016/j.eng.2018.11.030.

[12] B.E. Mneymneh, M. Abbas, H. Khoury, Automated hardhat detection for construction safety applications, Procedia Eng. 196 (2017) 895–902, https://doi.org/10.1016/j.proeng.2017.08.022.

[13] Kishor Shrestha, Pramen P. Shrestha, Dinesh Bajracharya, Evangelos A. Yfantis, Hard-hat detection for construction safety visualization, J. Constr. Eng. 2015 (1) (2015) 1–8, https://doi.org/10.1155/2015/721380.

[14] M. Kopsida, I. Brilakis, P. Vela, "A Review of Automated Construction Progress and Inspection Methods," Proceeding of the 32nd CIB W78 Conference 2015, 2015, pp. 421–431. http://itc.scix.net/cgi-bin/works/Show?w78-2015-paper-044.

[15] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, F. Peña-Mora, Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques, Autom. Constr. 20 (8) (2011) 1143–1155, https://doi.org/10.1016/j.autcon.2011.04.016.

[16] S.J. Prince, Computer Vision: Models, Learning, and Inference, Cambridge University Press, 2012. ISBN-13: 978–1107011793.

[17] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (May 2015) 436–444, https://doi.org/10.1038/nature14539. Nature Publishing Group.

[18] R. Szeliski, Computer Vision: Algorithms and Applications | Springer Science and Business Media, 2011, 978-1-84882-934-3.

[19] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, Kaori Togashi, Convolutional neural networks: an overview and application in radiology, Insights Imag. 9 (4) (2018) 611–629, https://doi.org/10.1007/s13244-018-0639-9.

[20] Xiaolei Ma, Zhuang Dai, Zhengbing He, Jihui Ma, Yong Wang, Yunpeng Wang, Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction, Sensors 17 (4) (2017) 818, https://doi.org/10.3390/s17040818.

[21] Yuanbin Wang, Pai Zheng, Xun Xu, Huayong Yang, Jun Zou, Production planning for cloud-based additive manufacturing—a computer vision-based approach, Robot. Comput. Integr. Manuf. 58 (2019) 145–157, https://doi.org/10.1016/j.rcim.2019.03.003.

[22] L. Matthies, M. Maimone, A. Johnson, Y. Cheng, R. Willson, C. Villalpando, A. Angelova, Computer vision on Mars, Int. J. Comput. Vis. 75 (1) (Oct. 2007) 67–92, https://doi.org/10.1007/s11263-007-0046-z.

[23] Ali Kazemian, Xiao Yuan, Omid Davtalab, Behrokh Khoshnevis, Computer vision for real-time extrusion quality monitoring and control in robotic construction, Autom. Constr. 101 (2019) 92–98, https://doi.org/10.1016/j.autcon.2019.01.022.

[24] Weili Fang, Lieyun Ding, Botao Zhong, Peter E.D. Love, Hanbin Luo, Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach, Adv. Eng. Inform. 37 (2018) 139–149, https://doi.org/10.1016/j.aei.2018.05.003.

[25] H. Seong, H. Choi, H. Cho, S. Lee, H. Son, C. Kim, Vision-Based Safety Vest Detection in a Construction Scene. ISARC.Proceedings of the International Symposium on Automation and Robotics in Construction 34, 2017, pp. 1–6, in: https://www.proquest.com/conference-papers-proceedings/vision-based-safety-vest-detection-construction/docview/1943517677/se-2?accountid=10920. accessed August 24, 2020.

[26] Lieyun Ding, Weili Fang, Hanbin Luo, Peter E.D. Love, Botao Zhong, Xi Ouyang, A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory, Autom. Constr. 86 (2018) 118–124, https://doi.org/10.1016/j.autcon.2017.11.002.

[27] W. Fang, B. Zhong, N. Zhao, P. Love E., H. Luo, J. Xue, S. Xu, A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network, Adv. Eng. Inform. 39 (2019) 170–177, https://doi.org/10.1016/j.aei.2018.12.005.

[28] Kevin K. Han, Mani Golparvar-Fard, Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs, Autom. Constr. 53 (2015) 44–57, https://doi.org/10.1016/j.autcon.2015.02.007.

[29] W. Fang, L. Ding, P. Love, E. Luo, H. Li, F. Pena-Mora, C. Zhou, Computer vision applications in construction safety assurance, Autom. Constr. 110 (2020) 103013, https://doi.org/10.1016/j.autcon.2019.103013.

[30] Pablo Martinez, Mohamed Al-Hussein, Rafiq Ahmad, A scientometric analysis and critical review of computer vision applications for construction, Autom. Constr. 107 (2019) 102947, https://doi.org/10.1016/j.autcon.2019.102947.

[31] B. Zhong, H. Wu, L. Ding, P. Love, H. Li, H. Luo, L. Jiao, Mapping computer vision research in construction: developments, knowledge gaps and implications for research, Autom. Constr. 107 (2019) 1029192019, https://doi.org/10.1016/j.autcon.2019.102919.

[32] A. Dionysiou, M. Agathocleous, C. Christodoulou, V. Promponas, Convolutional neural networks in combination with support vector machines for complex sequential data classification, Int. Conf. Artif. Neural Netw. (2018) 444–455, https://doi.org/10.1007/978-3-030-01421-6_43.

[33] Hanbin Luo, Chaohua Xiong, Weili Fang, Peter E.D. Love, Bowen Zhang, Xi Ouyang, Convolutional neural networks: computer vision-based workforce activity assessment in construction, Autom. Constr. 94 (2018) 282–289, https://doi.org/10.1016/j.autcon.2018.06.007.

[34] T.J. Yang, Y.H. Chen, V. Sze, Designing energy-efficient convolutional neural networks using energy-aware pruning, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2017) 5687–5695, https://doi.org/10.1109/CVPR.2017.643.

[35] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Proces. Syst. 25 (2012) 1097–1105, https://doi.org/10.1145/3065386.

[36] M. Pollefeys, D. Nistér, J. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, H. Towles, Detailed real-time urban 3D reconstruction from video, Int. J. Comput. Vis., Int. J. Comput. Vis. 78 (2) (2008) 143–167, https://doi.org/10.1007/s11263-007-0086-4.

[37] D. Shin, Z. Ren, E. Sudderth, C. Fowlkes, 3D scene reconstruction with multi-layer depth and epipolar transformers, Proc. IEEE/CVF Int. Conf. Comput. Vis. (2019) 2172–2182, https://doi.org/10.1109/ICCV.2019.00226.

[38] Tony Jebara, Ali Azarbayejani, Alex Pentland, 3D structure from 2D motion, IEEE Signal Process. Mag. 16 (3) (1999) 66–84, https://doi.org/10.1109/79.768574.

[39] Habib Fathi, Fei Dai, Manolis Lourakis, Automated as-built 3D reconstruction of civil infrastructure using computer vision: achievements, opportunities, and challenges, Adv. Eng. Inform. 29 (2) (2015) 149–161, https://doi.org/10.1016/j.aei.2015.01.012.

[40] Xian-Feng Han, Hamid Laga, Mohammed Bennamoun, Image-based 3D object reconstruction: state-of-the-art and trends in the deep learning era, IEEE Trans. Pattern Anal. Mach. Intell. 43 (5) (2019) 1578–1604, https://doi.org/10.1109/tpami.2019.2954885.

[41] Xiongwei Wu, Doyen Sahoo, Steven C.H. Hoi, Recent advances in deep learning for object detection, Neurocomputing 396 (2020) 39–64, https://doi.org/10.1016/j.neucom.2020.01.085.

[42] Waseem Rawat, Zenghui Wang, Deep convolutional neural networks for image classification: a comprehensive review, Neural Comput. 29 (9) (2017) 2352–2449, https://doi.org/10.1162/NECO_a_00990.

[43] Zhengxia Zou, Zhenwei Shi, Yuhong Guo, Jieping Ye, "Object Detection in 20 Years: A Survey." arXiv preprint arXiv:1905.05055. https://arxiv.org/pdf/1905.05055.pdf, 2019.

[44] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR) 01 (2005) 886–893, https://doi.org/10.1109/CVPR.2005.177.

[45] William S. Noble, What is a support vector machine? Nat. Biotechnol. 24 (12) (2006) 1565–1567, https://doi.org/10.1038/nbt1206-1565.

[46] Y. Cao, C. Wang, Z. Li, L. Zhang, L. Zhang, Spatial-bag-of-features, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2010) 3352–3359, https://doi.org/10.1109/CVPR.2010.5540021.

[47] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: a survey, Int. J. Comput. Vis. 128 (2) (2020) 261–318, https://doi.org/10.1007/s11263-019-01247-4.

[48] Xiangyuan Lan, Mang Ye, Shengping Zhang, Huiyu Zhou, Pong C. Yuen, Modality-correlation-aware sparse representation for RGB-infrared object tracking, Pattern Recogn. Lett. 130 (2020) 12–20, https://doi.org/10.1016/j.patrec.2018.10.002.

[49] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, T.K. Kim, Multiple object tracking: a literature review, Artif. Intell. 293 (Apr. 2021) 103448, https://doi.org/10.1016/j.artint.2020.103448. Elsevier B.V.

[50] Arnold W.M. Smeulders, Dung M. Chu, Rita Cucchiara, Simone Calderara, Afshin Dehghan, Mubarak Shah, Visual tracking: an experimental survey, IEEE Trans. Pattern Anal. Mach. Intell. 36 (7) (2013) 1442–1468, https://doi.org/10.1109/TPAMI.2013.230.

[51] A. Bewley, Z. Ge, L. Ott, F. Ramos, B. Upcroft, Simple online and realtime tracking, IEEE Int. Conf. Image Process. (ICIP) (2016) 3464–3468, https://doi.org/10.1109/ICIP.2016.7533003.

[52] D. Held, S. Thrun, S. Savarese, Learning to track at 100 FPS with deep regression networks BT - computer vision – ECCV 2016, Eur. Conf. Comput. Vis. (2016) 749–765, https://doi.org/10.1007/978-3-319-46448-0_45.

[53] H. Nam, B. Han, "Learning Multi-Domain Convolutional Neural Networks for Visual Tracking," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4293–4302. https://openaccess.thecvf.com/content_cvpr_2016/html/Nam_Learning_Multi-Domain_Convolutional_CVPR_2016_paper.html. date accessed September 17, 2020.

[54] F. Lateef, Y. Ruichek, Survey on semantic segmentation using deep learning techniques, Neurocomputing 338 (2019) 321–348, https://doi.org/10.1016/j.neucom.2019.02.003.

[55] Jiajia Luo, "Feature Extraction and Recognition for Human Action Recognition. " PhD diss, University of Tennessee, 2014. https://trace.tennessee.edu/utk_graddiss/2710. date accessed September 17, 2020.

[56] Lei Wang, Du Q. Huynh, Piotr Koniusz, A comparative review of recent kinect-based action recognition algorithms, IEEE Trans. Image Process. 29 (2019) 15–28, https://doi.org/10.1109/TIP.2019.2925285.
[57] K.K. Han, D. Cline, M. Golparvar-Fard, Formalized knowledge of construction sequencing for visual monitoring of work-in-progress via incomplete point clouds and low-LoD 4D BIMs, Adv. Eng. Inform. 29 (4) (2015), https://doi.org/10.1016/j.aei.2015.10.006.
[58] I. Brilakis, M.W. Park, G. Jog, Automated vision tracking of project related entities, Adv. Eng. Inform. 25 (4) (2011) 713–724, https://doi.org/10.1016/j.aei.2011.01.003.
[59] S. Shahandashti, S.N. Razavi, L. Soibelman, M. Berges, C. Caldas, I. Brilakis, Z. Zhu, Data-fusion approaches and applications for construction engineering, J. Constr. Eng. Manag.-ASCE 137 (10) (2011) 863–869, https://doi.org/10.1061/(asce)co.1943-7862.0000287.
[60] Bureau of Labor Statistics, Fatal Injury Report 2019 [Online]. Available: https://www.bls.gov/iif/oshwc/cfoi/cftb0324.htm, 2019. date accessed September 17, 2020.
[61] Abel Pinto, Isabel L. Nunes, Rita A. Ribeiro, Occupational risk assessment in construction industry–overview and reflection, Saf. Sci. 49 (5) (2011) 616–624, https://doi.org/10.1016/j.ssci.2011.01.003.
[62] H. Lingard, Occupational health and safety in the construction industry, Constr. Manag. Econ. 31 (6) (2013) 505–514, https://doi.org/10.1080/01446193.2013.816435.
[63] A. Bakri, R.M. Zin, M. Misnan, A. Mohammed, “Occupational Safety and Health (OSH) Management Systems: Towards Development of Safety and Health Culture,” 6th Asia-Pacific Structural Engineering and Construction Conference, no. September, 2006, pp. 5–6 [Online]. Available: http://eprints.utm.my/520/.
[64] S. Han, S. Lee, A vision-based motion capture and recognition framework for behavior-based safety management, Autom. Constr. 35 (2013) 131–141, https://doi.org/10.1016/j.autcon.2013.05.001.
[65] M. Wang, P. Wong, H. Luo, S. Kumar, V. Delhi, J. Cheng, “Predicting Safety Hazards among Construction Workers and Equipment using Computer Vision and Deep Learning Techniques,” Proceedings of the International Symposium on Automation and Robotics in Construction vol. 36, 2019, pp. 399–406, https://doi.org/10.22260/ISARC2019/0054.
[66] S. Chi, C.H. Caldas, Automated object identification using optical video cameras on construction sites, Comput.-Aided Civ. Infrastruct. Eng. 26 (5) (2011) 368–380, https://doi.org/10.1111/j.1467-8667.2010.00690.x.
[67] Qi Fang, Heng Li, Xiaochun Luo, Lieyun Ding, Timothy M. Rose, Wangpeng An, Yantao Yu, A deep learning-based method for detecting non-certified work on construction sites, Adv. Eng. Inform. 35 (2018) 56–68, https://doi.org/10.1016/j.aei.2018.01.001.
[68] Yantao Yu, Hongling Guo, Qinghua Ding, Heng Li, Martin Skitmore, An experimental study of real-time identification of construction workers’ unsafe behaviors, Autom. Constr. 82 (2017) 193–206, https://doi.org/10.1016/j.autcon.2017.05.002.
[69] Kinam Kim, Hongjo Kim, Hyoungkwan Kim, Image-based construction hazard avoidance system using augmented reality in wearable device, Autom. Constr. 83 (2017) 390–403, https://doi.org/10.1016/j.autcon.2017.06.014.
[70] S. Chi, C.H. Caldas, D.Y. Kim, A methodology for object identification and tracking in construction based on spatial modeling and image matching techniques, Comput.-Aided Civ. Infrastruct. Eng 24 (3) (2009) 199–211, https://doi.org/10.1111/j.1467-8667.2008.00580.x.
[71] Daeho Kim, Meiyin Liu, SangHyun Lee, Vineet R. Kamat, Remote proximity monitoring between mobile construction resources using camera-mounted UAVs, Autom. Constr. 99 (2019) 168–182, https://doi.org/10.1016/j.autcon.2018.12.014.
[72] Man-Woo Park, Nehad Elsafty, Zhenhua Zhu, Hardhat-wearing detection for enhancing on-site safety of construction workers, J. Constr. Eng. Manag. 141 (9) (2015), 04015024, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000974.
[73] Qi Fang, Heng Li, Xiaochun Luo, Lieyun Ding, Hanbin Luo, Chengqian Li, Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment, Autom. Constr. 93 (2018) 148–164, https://doi.org/10.1016/j.autcon.2018.05.022.
[74] Bahaa Eddine Mneymneh, Mohamad Abbas, Hiam Khoury, Vision-based framework for intelligent monitoring of hardhat wearing on construction sites, J. Comput. Civ. Eng. 33 (2) (2019) 04018066, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000813.
[75] Zdenek Kolar, Hainan Chen, Xiaowei Luo, Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images, Autom. Constr. 89 (2018) 58–70, https://doi.org/10.1016/j.autcon.2018.01.003.
[76] D. Roberts, T. Bretl, M. Golparvar-Fard, Detecting and classifying cranes using camera-equipped UAVs for monitoring Crane-Related Safety Hazards, Comput. Civ. Eng. 2017 (2017) 442–449, https://doi.org/10.1061/9780784480847.055.
[77] M. Neuhausen, J. Teizer, M. König, Construction Worker Detection and Tracking in Bird's-Eye View Camera Images. ISARC.Proceedings of the International Symposium on Automation and Robotics in Construction 35, 2018, pp. 1–8. https://www.proquest.com/docview/2123611289. accessed August 24, 2021.
[78] Z. Zhu, X. Ren, Z. Chen, Integrated detection and tracking of workforce and equipment from construction jobsite videos, Autom. Constr. 81 (2017) 161–171, https://doi.org/10.1016/j.autcon.2017.05.005.
[79] Madhuri Siddula, Fei Dai, Yanfang Ye, Jianping Fan, Classifying construction site photos for roof detection: a machine-learning method towards automated measurement of safety performance on roof sites, Constr. Innov. (2016), https://doi.org/10.1108/CI-10-2015-0052.
[80] Man-Woo Park, Ioannis Brilakis, Continuous localization of construction workers via integration of detection and tracking, Autom. Constr. 72 (2016) 129–142, https://doi.org/10.1016/j.autcon.2016.08.039.
[81] Weili Fang, Lieyun Ding, Hanbin Luo, Peter E.D. Love, Falls from heights: a computer vision-based approach for safety harness detection, Autom. Constr. 91 (2018) 53–61, https://doi.org/10.1016/j.autcon.2018.02.018.
[82] O. Golovina, J. Teizer, N. Pradhananga, Heat map generation for predictive safety planning: preventing struck-by and near miss interactions between workers-on-foot and construction equipment, Autom. Constr. 71 (2016) 99–115, https://doi.org/10.1016/j.autcon.2016.03.008.
[83] J. Teizer, P.A. Vela, Personnel tracking on construction sites using video cameras, Adv. Eng. Inform. 23 (4) (2009) 452–462, https://doi.org/10.1016/j.aei.2009.06.011.
[84] J. Yang, O. Arif, P.A. Vela, J. Teizer, Z. Shi, Tracking multiple workers on construction sites using video cameras, Adv. Eng. Inform. 24 (4) (2010) 428–434, https://doi.org/10.1016/j.aei.2010.06.008.
[85] M. Liu, S. Han, S. Lee, Tracking-Based 3D Human Skeleton Extraction from Stereo Video Camera toward an on-Site Safety and Ergonomic Analysis, Construction Innovation, 2016, https://doi.org/10.1108/CI-10-2015-0054.
[86] C. Li, S.H. Lee, Computer vision techniques for worker motion analysis to reduce musculoskeletal disorders in construction, Comput. Civ. Eng. 2011 (2011) 380–387, https://doi.org/10.1061/41182(416)47.
[87] S.J. Ray, J. Teizer, Real-time construction worker posture analysis for ergonomics training, Adv. Eng. Inform. 26 (2) (2012) 439–455, https://doi.org/10.1016/j.aei.2012.02.011.
[88] Idris Jeelani, Khashayar Asadi, Hariharan Ramshankar, Kevin Han, Alex Albert, Real-time vision-based worker localization & hazard detection for construction, Autom. Constr. 121 (2021) 103448, https://doi.org/10.1016/j.autcon.2020.103448.
[89] H. Ardiny, S. Witwicki, F. Mondada, “Construction Automation with Autonomous Mobile Robots: A review,” 3rd RSI International Conference on Robotics and Mechatronics (ICROM), 2015, pp. 418–424, https://doi.org/10.1109/ICRoM.2015.7367821.
[90] eTools : Construction eTool – Falls, Occupational Safety and Health Administration. https://www.osha.gov/etools/construction, 2021. Last Accessed 8/24/2021.
[91] GITHUB, OPENNI [Online]. Available: https://github.com/OpenNI/OpenNI, 2019. Last Accessed 8/24/2021.
[92] S. Han, S.H. Lee, F. Peña-Mora, Vision-based detection of unsafe actions of a construction worker: case study of ladder climbing, J. Comput. Civ. Eng. 27 (6) (2013) 635–644, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000279.
[93] Y.M. Ibrahim, Tim C. Lukins, X. Zhang, Emanuele Trucco, A.P. Kaka, Towards automated progress assessment of workpackage components in construction projects using computer vision, Adv. Eng. Inform. 23 (1) (2009) 93–103, https://doi.org/10.1016/j.aei.2008.07.002.
[94] A. Braun, S. Tuttas, U. Stilla, A. Borrmann, “Process- and Computer Vision-Based Detection of as-Built Components on Construction Sites,” Proceedings of the International Symposium on Automation and Robotics in Construction vol. 35, 2018, pp. 1–7, https://doi.org/10.22260/isarc2018/0091.
[95] S. Roh, Z. Aziz, F. Pena-Mora, An object-based 3D walk-through model for interior construction progress monitoring, Autom. Constr. 20 (1) (2011) 66–75, https://doi.org/10.1016/j.autcon.2010.07.003.
[96] Changmin Kim, Hyojoo Son, Changwan Kim, Fully automated registration of 3D data to a 3D CAD model for project progress monitoring, Autom. Constr. 35 (2013) 587–594, https://doi.org/10.1016/j.autcon.2013.01.005.
[97] Hyojoo Son, Changwan Kim, 3D structural component recognition and modeling method using color and 3D data for construction progress monitoring, Autom. Constr. 19 (7) (2010) 844–854, https://doi.org/10.1016/j.autcon.2010.03.003.
[98] H. Hamledari, B. McCabe, S. Davari, Automated Computer Vision-Based Detection of Components of under-Construction Indoor Partitions, Automation in Construction, 2017, https://doi.org/10.1016/j.autcon.2016.11.009.
[99] H. Kevin, D. Joseph, G.-F. Mani, Geometry- and appearance-based reasoning of construction progress monitoring, J. Constr. Eng. Manag. 144 (2) (Feb. 2018) 4017110, https://doi.org/10.1061/(ASCE)CO.1943-7862.0001428.
[100] T.C. Lukins, E. Trucco, Towards automated visual assessment of progress in construction project, in: Abhir Bhalerao, Nasir Rajpoot (Eds.), Proceedings of British Machine Vision Conference, 2007, pp. 1–10, https://doi.org/10.5244/C.21.18.
[101] Khashayar Asadi, Hariharan Ramshankar, Harish Pullagurla, Aishwarya Bhandare, Suraj Shanbhag, Pooja Mehta, Spondon Kundu, Kevin Han, Edgar Lobaton, Wu. Tianfu, Vision-based integrated mobile robotic system for real-time applications in construction, Autom. Constr. 96 (2018) 470–482, https://doi.org/10.1016/j.autcon.2018.10.009.
[102] Khashayar Asadi, Hariharan Ramshankar, Mojtaba Noghabaei, Kevin Han, Real-time image localization and registration with BIM using perspective alignment for indoor monitoring of construction, J. Comput. Civ. Eng. 33 (5) (2019), 04019031, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000847.
[103] F. Barbosa, J. Woetzel, J. Mischke, Reinventing Construction: A Route To Higher Productivity, McKinsey Global Institute, February 2017. http://dln.jaipuria.ac.in:8080/jspui/bitstream/123456789/2898/1/MGI-Reinventing-Construction-Full-report.pdf. Last Accessed 8/24/2021.
[104] J.Y. Kim, C.H. Caldas, Vision-based action recognition in the internal construction site using interactions between worker actions and construction objects, in: Proceedings of the International Symposium on Automation and Robotics in Construction vol. 30, 2013, p. 1, https://doi.org/10.22260/isarc2013/0072.

[105] D. Kashiwagi, J. Murphy, "A Contractor's Profit Analysis" Associated Schools of Construction (ASC) Proceedings of the 40th Annual Conference, April 8–10. http://ascpro0.ascweb.org/archives/cd/2004/2004pro/2003/Kashiwagi04a.htm, 2004.
[106] J. Gong, C.H. Caldas, An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations, Autom. Constr. 20 (8) (2011) 1211–1226, https://doi.org/10.1016/j.autcon.2011.05.005.
[107] Jie Gong, Carlos H. Caldas, Computer vision-based video interpretation model for automated productivity analysis of construction operations, J. Comput. Civ. Eng. 24 (3) (2010) 252–263, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000027.
[108] Jun Yang, Patricio Vela, Jochen Teizer, Zhongke Shi, Vision-based tower crane tracking for understanding construction activity, J. Comput. Civ. Eng. 28 (1) (2014) 103–112, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000242.
[109] Jochen Teizer, Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites, Adv. Eng. Inform. 29 (2) (2015) 225–238, https://doi.org/10.1016/j.aei.2015.03.006.
[110] E.R. Azar, B. McCabe, Automated visual recognition of dump trucks in construction videos, J. Comput. Civ. Eng. 26 (6) (2012) 769–781, https://doi.org/10.1061/(asce)cp.1943-5487.0000179.
[111] K.K. Han, M. Golparvar-Fard, Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs, Autom. Constr. 53 (2015), https://doi.org/10.1016/j.autcon.2015.02.007.
[112] J. Yang, Z. Shi, Z. Wu, Vision-Based Action Recognition of Construction Workers Using Dense Trajectories, Advanced Engineering Informatics, 2016, https://doi.org/10.1016/j.aei.2016.04.009.
[113] Jie Gong, Carlos H. Caldas, A computer vision and reasoning based video interpretation system for rapid productivity analysis of construction operations, in: 17th International Workshop on Intelligent Computing in Engineering, EG-ICE 2010. Nottingham, 2019. ISBN (Electronic) 9781907284601, https://www.researchwithrutgers.com/en/publications/a-computer-vision-and-reasoning-based-video-interpretation-system.
[114] J. Gong, C.H. Caldas, C. Gordon, Learning and classifying actions of construction workers and equipment using bag-of-video-feature-words and Bayesian network models, Adv. Eng. Inform. 25 (4) (2011) 771–782, https://doi.org/10.1016/j.aei.2011.06.002.
[115] Ali Behnam, Darshana Charitha Wickramasinghe, Moataz Ahmed Abdel Ghaffar, Vu Tuong Thuy, Yu Hoe Tang, Hazril Bin Md Isa, Automated progress monitoring system for linear infrastructure projects using satellite remote sensing, Autom. Constr. 68 (2016) 114–127, https://doi.org/10.1016/j.autcon.2016.05.002.
[116] E.R. Azar, B. McCabe, Part based model and spatial-temporal reasoning to recognize hydraulic excavators in construction images and videos, Autom. Constr. 24 (2012) 194–202, https://doi.org/10.1016/j.autcon.2012.03.003.
[117] M. Memarzadeh, M. Golparvar-Fard, J. Carlos Niebles, Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors, Autom. Constr. 32 (2013) 24–37, https://doi.org/10.1016/j.autcon.2012.12.002.
[118] Mani Golparvar-Fard, Arsalan Heydarian, Juan Carlos Niebles, Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers, Adv. Eng. Inform. 27 (4) (2013) 652–663, https://doi.org/10.1016/j.aei.2013.09.001.
[119] Jie Gong, Carlos H. Caldas, An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations, Autom. Constr. 20 (8) (2011) 1211–1226, https://doi.org/10.1016/j.autcon.2011.05.005.
[120] Jie Gong, Carlos H. Caldas, An intelligent video computing method for automated productivity analysis of cyclic construction operations, Comput. Civ. Eng. 2009 (2009) 64–73, https://doi.org/10.1061/41052(346)7.
[121] Bon-Gang Hwang, Stephen R. Thomas, Carl T. Haas, Carlos H. Caldas, Measuring the impact of rework on construction cost performance, J. Constr. Eng. Manag. 135 (3) (2009) 187–198, https://doi.org/10.1061/(ASCE)0733-9364(2009)135:3(187).
[122] Pingbo Tang, Daniel Huber, Burcu Akinci, Robert Lipman, Alan Lytle, Automatic reconstruction of as-built building information models from laser-scanned point clouds: a review of related techniques, Autom. Constr. 19 (7) (2010) 829–843, https://doi.org/10.1016/j.autcon.2010.06.007.
[123] Kuo-Liang Lin, Jhih-Long Fang, Applications of computer vision on tile alignment inspection, Autom. Constr. 35 (2013) 562–567, https://doi.org/10.1016/j.autcon.2013.01.009.
[124] Reza Maalek, Derek D. Lichti, Janaka Y. Ruwanpura, Automatic recognition of common structural elements from point clouds for automated progress monitoring and dimensional quality control in reinforced concrete construction, Remote Sens. 11 (9) (2019) 1102, https://doi.org/10.3390/rs11091102.
[125] Andrey Dimitrov, Mani Golparvar-Fard, Segmentation of building point cloud models including detailed architectural/structural features and MEP systems, Autom. Constr. 51 (2015) 32–45, https://doi.org/10.1016/j.autcon.2014.12.015.
[126] Fu Jing, "A Strong Start for 2020," National Association of Home Builders Discusses Economics and Housing Policy, Feb 7. https://eyeonhousing.org/2020/02/a-strong-start-for-2020/, 2020. Last Accessed 8/24/2021.
[127] A.V. Malakhov, D.V. Shutin, S.G. Popov, Bricklaying robot moving algorithms at a construction site, IOP Conf. Ser. : Mater. Sci. Eng. 734 (2020), 012126, https://doi.org/10.1088/1757-899X/734/1/012126.
[128] Khaled Elashry, Ruairi Glynn, An approach to automated construction using adaptive programing, in: Robotic Fabrication in Architecture, Art and Design 2014, Springer, Cham, 2014, pp. 51–66, https://doi.org/10.1007/978-3-319-04663-1_4.
[129] A. Bulat, G. Tzimiropoulos, How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks), Proc. IEEE Int. Conf. Comput. Vis. (2017) 1021–1030, https://doi.org/10.1109/ICCV.2017.116.
[130] M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, Int. Conf. Mach. Learn. (2019) 6105–6114. https://arxiv.org/abs/1905.11946.
[131] T. Tommasi, N. Patricia, B. Caputo, T. Tuytelaars, A deeper look at dataset bias, in: Domain Adaptation in Computer Vision Applications, Springer, 2017, pp. 37–55, https://doi.org/10.1007/978-3-319-58347-1_2.
[132] K.K. Han, M. Golparvar-Fard, Multi-sample image-based material recognition and formalized sequencing knowledge for operation-level construction progress monitoring, Comput. Civ. Build. Eng. (2014) 364–372, https://doi.org/10.1061/9780784413616.046.
[133] Y. Jung, S. Woo, Flexible work breakdown structure for integrated cost and schedule control, J. Constr. Eng. Manag. (2004), https://doi.org/10.1061/(ASCE)0733-9364(2004)130:5(616).
[134] JoonOh Seo, SangUk Han, SangHyun Lee, Hyoungkwan Kim, Computer vision techniques for construction safety and health monitoring, Adv. Eng. Inform. 29 (2) (2015) 239–251, https://doi.org/10.1016/j.aei.2015.02.001.
[135] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, Li Fei-Fei, ImageNet: a large-scale hierarchical image database, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248–255, https://doi.org/10.1109/CVPR.2009.5206848.
[136] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, C. Lawrence Zitnick, Microsoft COCO: Common objects in context, in: Lecture Notes in Computer Science vol. 8693, 2014, pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48. LNCS, no. PART 5.
[137] S. Xin, S. Nousias, K.N. Kutulakos, A.C. Sankaranarayanan, S.G. Narasimhan, I. Gkioulekas, A theory of fermat paths for non-line-of-sight shape reconstruction, Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (2019) 6800–6809, https://doi.org/10.1109/CVPR.2019.00696.