Front. Struct. Civ. Eng. 2025, 19(6): 1021–1040
https://doi.org/10.1007/s11709-025-1197-0

RESEARCH ARTICLE

Web-based multi-vision platform for earthwork productivity on construction sites using real-time model updating

Jeongbin HWANG[a], Insoo JEONG[b,c], Junghoon KIM[b,c], Seokho CHI[c,d*]

[a] *Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC 27695, USA*
[b] *Site Vision Incorporated, Seoul 08826, South Korea*
[c] *Department of Civil and Environmental Engineering, Seoul National University, Seoul 08826, South Korea*
[d] *Institute of Construction and Environmental Engineering, Seoul National University, Seoul 08826, South Korea*
[*] *Corresponding author. E-mail: shchi@snu.ac.kr*

ABSTRACT Earthwork productivity analysis is essential for successful construction projects. If productivity analysis results can be accessed anytime and anywhere, then project management can be performed more efficiently. To this end, this paper proposes an earthwork productivity monitoring framework via a real-time scene updating multi-vision platform. The framework consists of four main processes: 1) site-optimized database development; 2) real-time monitoring model updating; 3) multi-vision productivity monitoring; and 4) web-based monitoring platform for Internet-connected devices. The experimental results demonstrated satisfactory performance, with an average macro F1-score of 87.3% for continuous site-specific monitoring, an average accuracy of 86.2% for activity recognition, and the successful operation of multi-vision productivity monitoring through a web-based platform in real time. The findings can contribute to supporting site managers to understand real-time earthmoving operations while achieving better construction project and information management.

KEYWORDS online-active learning, site-customized monitoring, multi-vision monitoring, earthwork productivity analysis, web-based site monitoring platform

1 Introduction

The construction industry is a key sector impacting national gross domestic product, economic growth, and employment [1–3]. Enhancing productivity in construction projects can offer significant benefits both nationally and corporately. However, productivity growth in the construction industry is slower compared to other industries, with global construction productivity growing at 1.0% per year, compared to 2.7% for the global economy and 3.6% for manufacturing [4,5]. Many researchers are thus actively endeavoring to improve construction productivity, particularly in earthwork, which accounts for about 20% of total construction project costs [6,7].

Article history: Received Sep 4, 2024; Accepted Mar 20, 2025

Productivity monitoring in construction involves observing the efficiency of resources on-site, identifying factors inhibiting productivity, and making necessary adjustments [8–11]. This process is essential for successful construction project management and can support various decision-makings made by site managers. Recently, productivity monitoring has been attempted through various methods, including site visits by managers, or using diverse sensors, such as radio-frequency identification, global positioning system, and ultra-wide band [12–14]. Among these various approaches, vision-based productivity monitoring based on computer vision technology is being extensively researched for practical application in the field due to its easy installation and capacity to explain site conditions [15–18]. Cameras can collect more than just data on the type and location of target construction resources; they

can also gather video stream data about the overall scene information where construction operation is taking. Hence, many researchers are endeavoring hard to apply such computer vision technology for vision-based productivity monitoring. This approach not only easily analyzes the location and behavior information of each construction resource but also offers the advantage of analyzing additional information including environmental conditions and information of non-target construction resources.

Most construction projects often span several years, requiring long-term productivity monitoring under ever changing site conditions. For successful site management, it is essential to be able to access productivity monitoring results anytime and anywhere. However, vision-based monitoring models are trained on images previously taken and gathered at distinct time intervals. As the construction project progresses, the visual features of construction sites (e.g., types of operating heavy equipment) change, which can lead to a drop in the model's performance. Given this, continuous high-performance monitoring requires the development of a large, high-quality site-specific training image Database (DB) to retrain the model periodically when the model's performance decreases. Furthermore, it is more efficient to understand widespread on-site situations when recording and collecting video data from multiple cameras simultaneously rather than relying on just one. However, since videos captured by multiple cameras are collected separately, an additional process is required for integrated information management. For comprehensive productivity monitoring of the site, if an object appears in one video and then disappears, only to be captured again in another video, the system must recognize that they are the same object and integrate the productivity-related information obtained from both videos.

To overcome the existing challenges, the authors propose an earthwork productivity monitoring framework via a real-time scene updating platform with multiple videos that are recorded by different cameras. The proposed framework consists of four main processes: 1) site-optimized DB development; 2) real-time monitoring model updating; 3) multi-vision productivity monitoring; and 4) web-based monitoring platform for Internet-connected devices. The authors developed a multi-vision productivity monitoring framework for construction equipment that can perform various earthmoving operations. This approach allows for the rapid development of site-optimized training image DB and ensures that the monitoring model can recognize when its performance decreases, minimizing user effort. It also integrates vision-based monitoring results from multiple videos and statistically analyzes productivity to automatically detect and propose solutions for productivity inhibitors.

This research has the following contributions. First, the proposed framework can be applied to any other construction sites without relying on viewpoints and target resources. Second, the vision-based monitoring model can be updated in real time for dynamic construction sites. Third, the authors integrated vision-based monitoring results from multiple videos by considering the conditions of the construction site. Fourth, real-time monitoring of large and complex construction sites has become possible, and the results can be monitored from anywhere at any time. Last, this research achieves the sustainable, high-performed, and long-term site monitoring. The authors focused on heavy equipment (i.e., dump trucks, excavators, dozers, and rollers) that perform various operations in adverse working environments at multiple zones. After this introduction, the paper reviews previous research on vision-based productivity monitoring. It then explains the processes involved in the framework and presents experimental results to validate the framework. Finally, the conclusions drawn from the research are discussed.

2 Related work

Numerous researchers endeavored to monitor construction sites automatically using deep learning-based computer vision techniques. Object detection is one of the most basic algorithms, and comprehensive efforts have been made to apply it to construction sites. For example, Kim et al. [19] tested the feasibility of a Faster Region-based Convolutional Neural Network (Faster R-CNN) [20] with Residual Neural Network-101 (ResNet-101) [21] to detect various types of construction resources, including workers, heavy equipment, and materials. Similarly, Li et al. [22] and Shin et al. [23] adopted You Only Look Once (YOLO) to detect resources (e.g., rebars and heavy equipment) on construction sites. Nath and Behzadan [24] also applied YOLO on construction sites under different visual conditions. Many other studies have been actively conducted on object tracking. Object tracking incorporates the temporal concept of consecutive images to determine whether objects in previous and subsequent frames are the same or different. For instance, Zhu et al. [25] combined a detector and tracker to leverage their strengths and mitigate their weaknesses in tracking construction resources. Xiao and Zhu [26] compared 15 tracking methods on sites and indicated that those built on sparse representation were more effective than the others. There were further attempts to increase the performance of object tracking. Jeong et al. [27] enhanced a tracking algorithm by unsupervised clustering-driven post-processing for curtain wall installation. Xiao and Kang [28] proposed a construction machine tracker, which detects construction machines by YOLOv3 in each frame and connects the detection results

of two consecutive frames for tracking. Angah and Chen [29] proposed a gradient-based method for tracking multiple workers on construction sites. They improved the performance through detecting, matching, and re-matching the workers. Yan et al. [30] tracked dump trucks to monitor material arrival delays by integrating a deep CNN, a Kanade-Lucas-Tomasi corner feature tracker, and a hash-based occlusion handling strategy. These have proven the feasibility of vision-based tracking methods to monitor productivity.

The findings of earlier works led researchers to apply vision-based algorithms to recognize construction activities. Luo et al. [31] used three different CNNs that analyzed red-green-blue, optical flow, and gray channels for recognizing construction activities. Kim and Chi [32] and Zhang et al. [33] combined CNN and long short-term memory to generate time-spatial and temporal features of equipment operations. In other studies, Luo et al. [34] integrated relevance networks on CNN to extract semantic relevance representing the likelihood of cooperation or coexistence between different resources for activity recognition. Similarly, Roberts and Golparvar-Fard [35] used a hidden Markov model to understand the temporal sequence of earthmoving activities, and Soltani et al. [36] applied stereo-vision for estimating three-dimensional poses of excavators.

Recently, the majority of construction sites are equipped with multiple Closed-Circuit Television (CCTV) cameras, enabling comprehensive video monitoring of every working spot of the site. However, there is a limitation with vision-based monitoring to analyze multiple scene videos captured simultaneously. Unlike humans, when a resource captured by one CCTV appears in the video of another CCTV, the monitoring model shows difficulties from inherently determining if the two resources are the same one. Several studies have been conducted to empower vision-based monitoring to discern this information automatically. For example, Cheng et al. [37] designed a similarity loss function to encourage deep learning models to learn discriminative human features for robust tracking of individual workers. Zhang et al. [38] also applied the re-identification algorithm of workers, which adopts the distance-metric-based deep model. Wei et al. [39] extracted the feature maps using a spatial attention network. Existing computer vision techniques exhibit high performance only for images with visual characteristics similar to the train data. However, as construction project progresses over time, the visual characteristics of the site background change, leading to a gradual degradation in the performance of vision-based models. Additionally, to monitor a construction site successfully, it is necessary to analyze videos collected from multiple cameras simultaneously and to track each construction resource operated on diverse spots. For this, an optimized vision-based monitoring model is required for each camera, and these models must be capable of real-time updates. This paper resolved this by applying a baseline DB and an online-active learning approach, thereby implementing real-time updating of vision-based models.

Lastly, construction productivity can be defined in various ways, such as the amount of work completed per hour, or the time and cost required for a specific task. While its definition may vary depending on context, productivity is generally considered high when a task is completed in less time or at a lower cost. Most studies analyzing site productivity through vision-based monitoring estimate productivity based on task duration, defining it as higher when less time is required. Accordingly, productivity analysis is conducted by estimating the time each resource spends on a task. Kim et al. [40] estimated productivity by using CNN-based equipment detection results as inputs to an earthmoving simulation model. Besides, Kim et al. [41] applied deep neural network-based equipment and license plate detection results to track excavators on multiple videos and estimate earthwork productivity. Cheng et al. [42] conducted an excavator's action recognition to calculate both work time and average cycle time for productivity analysis. However, these studies calculated the productivity based on the simulation while not reflecting the dynamic and complex nature of construction sites. To address the limitation, Bügler et al. [43] analyzed earthwork productivity by not only employing activity recognition but also using photogrammetry to create a point cloud to measure the volume of excavated soil. Additionally, Chen et al. [44] applied zero-shot learning for activity recognition to consider the characteristics of construction sites dynamically changing. However, from the practical application point of view, there is still a lack of studies that detect productivity from the multiple viewpoints to cover the entire site and identify productivity inhibitors and their causes. Furthermore, for successful site management, managers need to be able to confirm productivity analysis results anytime and anywhere. This study developed a web-based multi-vision platform accessible from any device, such as a smartphone, tablet personal computer and laptop.

3 Proposed framework

Figure 1 illustrates the proposed framework that includes four main processes: 1) site-optimized DB development; 2) real-time monitoring model updating; 3) multi-vision productivity monitoring; and 4) web-based monitoring platform for Internet-connected devices. The details of these processes are described in the following subsections.

3.1 Site-optimized database development

Due to the diverse visual characteristics present at

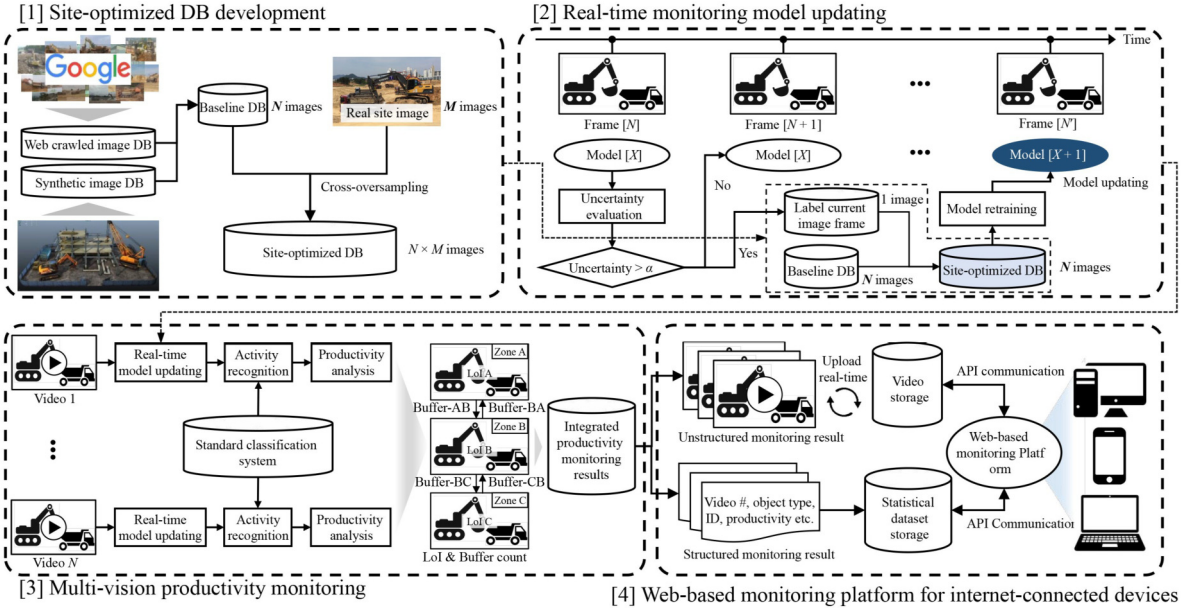

Fig. 1 Flowchart of the proposed framework.

individual construction sites, such as the backgrounds of the sites and the types of existing resources, it is necessary to generate a new training image DB to train a vision-based monitoring model. Constructing a large and high-quality training image DB, however, requires substantial effort. It involves visiting the site in person to install cameras, collecting video stream data, and annotating construction resources. During earthwork operations, the background of the site image continuously changes while the types of resources involved on the site remain consistent. Taking this into account, the authors aim to develop a DB that considers only the changes in the background of the construction site. We utilize real construction site images as the background for the images and develop a site-optimized DB by combining these backgrounds with a pre-generated image DB of construction resources.

To this end, it is beneficial to pre-develop a DB that embodies the visual characteristics of construction resources, named a baseline DB. This research automatically develops and collects a baseline DB using web crawling [45] and virtual reality techniques [46] with the application of the authors' previous research findings. After that, the authors adopted a site-optimized DB [47], a large and high-quality image DB, by cross-oversampling the baseline DB and the previous real target site image. In detail, the resources only change in location or pose, while their other visual characteristics remain unchanged. On the other hand, the site background continuously changes as the construction process progresses. Considering this, the construction resources in the pre-developed baseline DB were resized and synthesized into real site images. The synthesized areas were placed in appropriate locations (e.g., an excavator cannot be in the sky, so it was only synthesized on the ground). Hwang et al. [47] demonstrated that a model trained using this approach can avoid background mismatches. Utilizing the site-optimized DB enables the development of a high-performance vision-based model with only a small number of target site images. Furthermore, since the baseline DB is already established, the time required for data preparation in actual field environments can be minimized.

3.2 Real-time monitoring model updating

Maintaining the high performance of vision-based monitoring models requires a large and high-quality training image DB and the model retraining to improve its performance whenever it decreases. To determine the performance decreases in real time, practitioners must monitor the analysis results and compare them with the ground truths. However, this traditional method demands the same effort as manually monitoring the site, making the use of computer vision techniques meaningless. To solve this problem, the model should be able to determine when its performance decreases and update itself in real time. To achieve this, the authors utilized the concept of active and online learning approaches. Active learning [48] is a semi-supervised learning technique that selects the most meaningful data from a large amount of training data and incrementally maximizes the performance of the analysis model by learning from it. Specifically, the uncertainty of the model's predictions for each data point is evaluated, and data with high uncertainty is prioritized for learning [49,50]. The ability of the active learning approach to evaluate uncertainty and determine the need for learning is essential for maintaining model performance [51,52]. Second, online learning is another semi-unsupervised learning technique [53] that generates

training data from actual test data to optimize model training in the test environment. It is suitable for analyzing video stream data because it processes and learns data sequentially. Relying solely on online learning does not allow the vision-based models to determine when training is required.

To solve these drawbacks, the authors propose an online-active learning algorithm for real-time monitoring model updating. The model analyzes the real-time video frames currently and evaluates the uncertainty. The model is maintained if the uncertainty does not exceed the threshold. If the uncertainty exceeds the threshold, it is determined that the model needs to be updated. The uncertainty is defined as the average confidence score of the last ten frames. The performance evaluation of object detection models is commonly conducted by comparing the predicted values with a confidence score of 0.5 or higher to the ground truth values. A confidence score of 0.5 or higher means that there is more than a 50% probability of being correct, and thus, the model considers these predictions as "correct". This is why this threshold is chosen. Therefore, only confidence scores of 0.5 or higher are used for calculating uncertainty. When evaluating the performance of a vision-based model, the predicted values of the model are compared with ground truth values that are manually generated by humans. However, in online active learning, the model's performance degradation must be assessed in real time, making it impossible to generate ground truth values. As a result, the existing evaluation method cannot be used to determine the need for model updates. Instead, uncertainty is defined based on the confidence scores of the model results, which is then used to predict model performance and decide whether an update is necessary (Eq. (1)). Consequently, lowering the uncertainty threshold increases the confidence score threshold, causing the online active learning loop to operate more frequently and improving model performance. In statistical analysis, confidence intervals are commonly used at levels of 0.9 or higher, reflecting a general acceptance of 0.9 as a high standard. Furthermore, the authors set a high threshold of 0.9 to ensure the performance of the model because the framework must rely solely on the predictions generated by itself. For this reason, the uncertainty threshold was set at 0.1, demonstrating sufficient training frequency and performance in this paper. This study applied YOLOv5 [54] with 10 epochs for the object detection module, which is one of the powerful object detection algorithms. When the model needs to be updated, a training image DB must be required. Therefore, the authors applied the large and high-quality site-optimized DB which was mentioned previously.

$$Uncertainty = 1 - \mathrm{avg}\left(\sum_{i}^{10} Confidence\ Score_i\right). \quad (1)$$

3.3 Multi-vision productivity monitoring

To monitor earthwork productivity at construction sites, it is necessary to recognize the activity of resources in image frames, single-vision productivity analysis, and integrate the multiple single-vision productivity analysis results. For this purpose, this section covers action recognition, single-vision productivity analysis, and multiple productivity analysis integration. For action recognition, the object tracking algorithm is used to assign IDs to each construction resource. The research applies the centroid tracker that tracks the centroid of the bounding boxes moving and changing over time. Based on the tracking results, the authors identified individual actions and interactions to recognize activities (Fig. 2). First, the individual actions of each construction resource are determined. Individual actions refer to movements do not assign any meaning, and categorize them into move, stop, and semi-stop. "Move" refers to changing location, "semi-stop" represents movements in place but without changing the location, and "stop" refers to a state of no movement. The individual actions of move, semi-stop, and stop are determined based on the presence or absence of movement of the construction resource's centroid and changes in the aspect ratio of the bounding box. When the resource's centroid moves, it is classified as a "move". When the resource's centroid does not move, but the aspect ratio changes, it is determined as "semi-stop". When neither the object's centroid moves nor the aspect ratio changes, it is classified as "stop". Specifically, the centroid movement is calculated as the ratio between the change in the center of the bounding box and the length of its diagonal. The presence of any change in the aspect ratio is determined using the aspect ratio change rate, with the threshold set to 0.1 in one second for all measurements. Second, based on the distance and actions of resources, interaction between the resources is determined. For example, loading work refers to the act of an excavator loading soil onto a dump truck. In this case, a dump truck is near the excavator, and the excavator continuously moves its arm and body in a "semi-stop" state. From the dump truck's perspective, the excavator is nearby and will be in a "stop" state. Since the activities vary depending on the type of construction resource, it is necessary to set rules for each resource to be analyzed. The presence of interaction is determined by comparing the distance between the centers of the bounding boxes with half the sum of their diagonals. The activities can be classified into three different production states: productive, semi-productive, and non-productive. Productive activities directly impact productivity (e.g., load), semi-productive activities indirectly affect productivity (e.g., travel), and non-productive activities have no impact on productivity (e.g., idle). This research calls this a standard classification system (Table 1). For

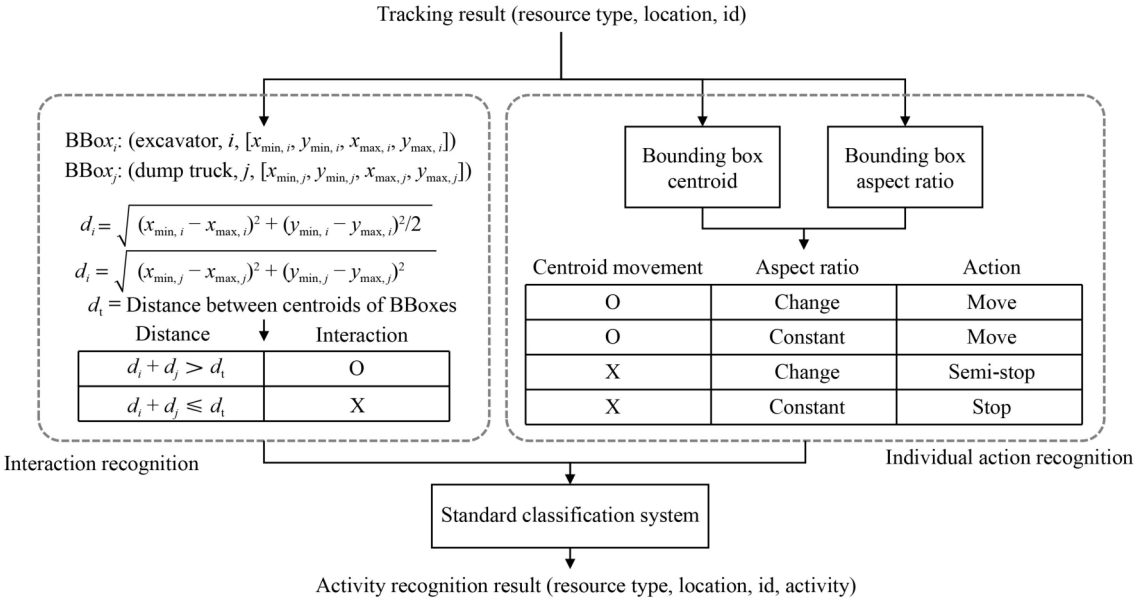

Fig. 2 Process of the activity recognition using interaction and individual action recognition.

Table 1 Standard classification system

Resource type	Production state	Activity type	Individual action	Interaction	
				Construction resource	State
Excavator	non-productive	idle	stop	–	–
	semi-productive	travel	move	–	–
		soil-manage	semi-stop	–	–
	productive	load	semi-stop	dump truck	stop
Dump truck	non-productive	idle	stop	–	–
	semi-productive	travel	move	–	–
	productive	load	stop	excavator	semi-stop
		unload	semi-stop	–	–
Dozer	non-productive	idle	stop	–	–
	productive	soil-manage	move	–	–
Roller	non-productive	idle	stop	–	–
	productive	soil-manage	move	–	–

example, the activity types of excavators are categorized as idle, travel, soil-manage, and load. Notably, during the loading operation, there is an interaction with a dump truck in a stop state. Conversely, the excavator is in a "semi-stop" state during the soil-managing operation, such as a loading operation. The soil-managing operation but does not have any interaction with other construction resources. Additionally, the excavator exhibits individual actions such as stop, move, and semi-stop while performing each operation. The semi-stop action refers to the state where the excavator moves its arm or body at the exact location for soil management or loading (Fig. 3).

Earthwork productivity can be defined in a variety of ways. Among them, the authors defined earthwork productivity as (a) the total earthwork volume and (b) the productive work time ratio of each resource. The total earthwork volume is calculated by multiplying the number of loaded dump trucks by their capacity. The productive work time ratio quantifies the activity time duration each resource spends on various tasks. For single-vision productivity analysis, the productive work time of each resource on each video is comparatively analyzed. The productive work time ratio is defined as the proportion of 'productive' (load and unload) and 'semi-productive' (travel) activities within the total working time. Dump trucks with significantly longer durations for travel, load, or unload activities are scrutinized, as such patterns could indicate reduced productivity. These resources are considered inhibitor factors, and statistical analysis using Z-score normalization is conducted to

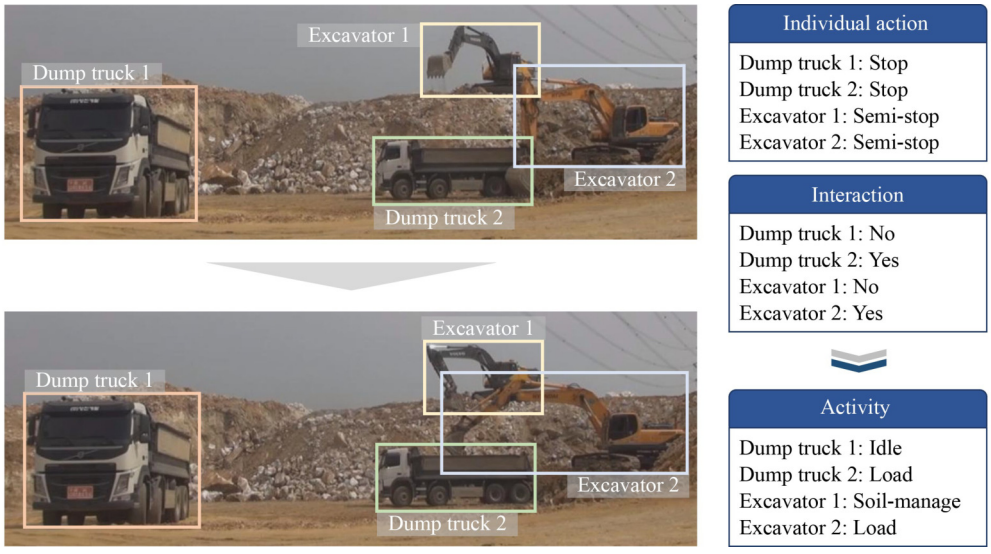

Fig. 3 Example of activity recognition via the standard classification system.

detect them. Resources with a *Z*-score higher than the confidence interval (95%, *Z*-score threshold: 1.96) are deemed as outliers and they are assumed as inhibitors.

Multi-vision productivity monitoring requires integration of multiple productivity analysis results. The authors applied rule-based object re-identification which consists of the Line of Interest (LoI) and buffer algorithm [55]. The LoI method determines whether a resource has left the field of view, while the buffer algorithm stores the type and ID of resources transitioning between zones. The LoI involves creating a virtual line in the captured video (Fig. 4). If a construction resource crosses that line, such as moving out of the video frame, it is determined that the resource has exited the zone. Conversely, if a resource crosses the virtual line in the opposite direction, it is concluded that the resource has entered the zone. Utilizing this principle, if the last centroid coordinate of a tracked resource crosses the virtual line, it means that the object has left the site and is moving to another location. The virtual lines are set differently based on the characteristics of the viewpoints. Once the camera is installed, the recording zone remains essentially unchanged, ensuring that these virtual lines stay fixed until filming concludes.

The buffers are defined to serve as links between actual zones. They function as an algorithm that temporarily stores the types and IDs of target construction resources while it transitions from exiting one zone to entering another. This helps preserve the IDs of resources that disappear from all video frames and aids in predicting which area (blind spot) the resource is located in during the transition (Fig. 5). For instance, when a construction resource exits zone A, the buffer-AB stores the resource. After a while, if a construction resource enters zone B from zone A, which has the same type as the resource stored in buffer-AB, buffer-AB removes the stored resource and takes the ID for object re-identification. In detail, the minimum and maximum duration of using the buffer is set, because the resources need minimum time to travel and reduce tracking errors by refreshing the buffer.

3.4 Web-based monitoring platform for Internet-connected devices

If the results of site monitoring can be observed from anywhere at any time, site managers can make more proactive decisions for successful site management. This section is for developing a web-based monitoring platform to display the multi-vision monitoring results on various Internet-connected devices without spatiotemporal constraints. Considering this, a web-based platform user interface is developed to display both real-time multi-vision monitoring and productivity monitoring results (Fig. 6). The platform is developed using Amazon Web Service (AWS) Elastic Compute Cloud (EC2) for multi-vision monitoring, AWS Simple Storage Service (S3) for storing results, and AWS Lightsail for the other basic functions.

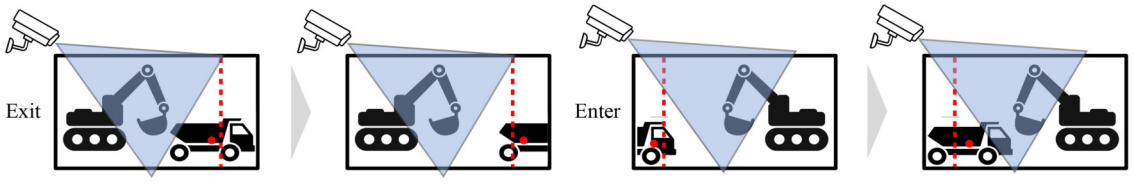

Fig. 4 Example of LoI use (Left: exit; Right: enter).

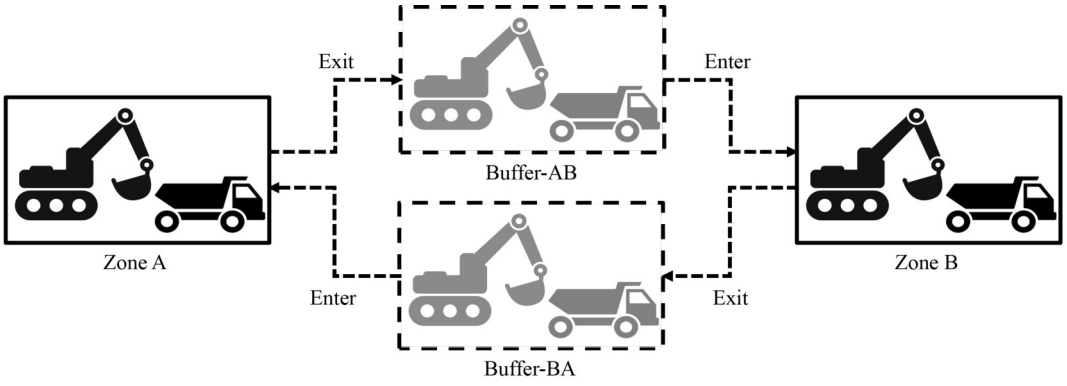

Fig. 5 Actual zones and buffers for object re-identification.

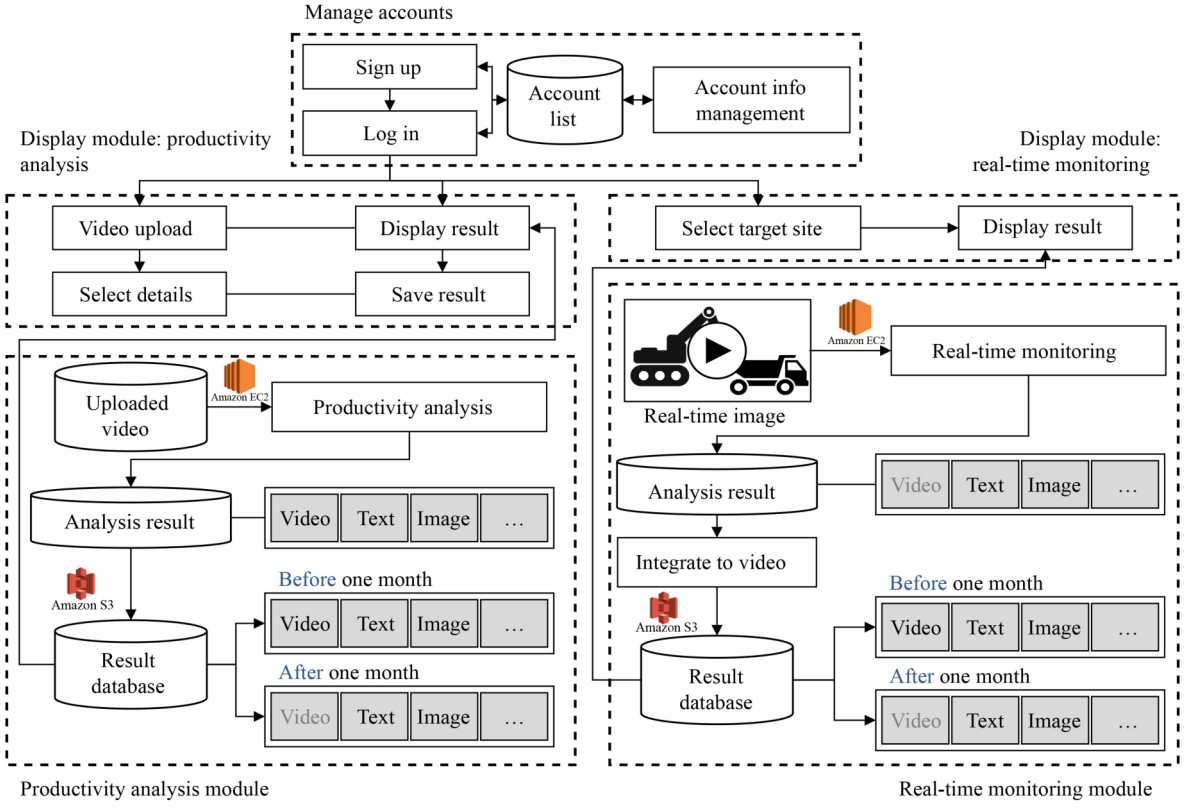

Fig. 6 Real-time monitoring model updating process.

For the real-time multi-vision monitoring results, the following processes must be performed in real time: (a) downloading each frame from the video stream data; (b) analyzing each frame without delay; (c) uploading the analysis result to the back-end server; and (d) converting the analyzed frames into video stream data for platform users. To do process (b) in real time, different types of AWS EC2 instances were validated for their suitability with object detection algorithms such as Faster R-CNN and YOLOv5 (Table 2). According to Hwang et al. [47], the performance difference between Faster R-CNN and YOLOv5 is minimal when trained with a site-optimized DB. Therefore, the model selection was based on analysis time and instance cost. The most affordable option, g4dn.xlarge with YOLOv5, was chosen for its balance between analyzing time and cost. To perform processes (a) and (c) in real time, AWS S3 buckets are used for storing numerous analyzed image frames and serving as the platform's back-end server. Three separate codes operate on the back-end of the AWS EC2 instance to minimize delays in downloading, analyzing, and uploading. The other one for displaying image frames on the platform operates on the back-end of the AWS Lightsail instance.

The user interface utilizes uploaded video data for productivity monitoring. The productivity monitoring results are uploaded and displayed automatically for the users. For real-time multi-vision monitoring, the system

Table 2 Analysis time and cost according to AWS EC2 instance and network type

Instance type	Faster R-CNN (s/frame)	YOLOv5 (s/frame)	Instance cost (USD/h)
p3.2xlarge	0.39	0.02	3.060
p3.16xlarge	0.39	0.02	24.480
p2.xlarge	0.87	0.04	0.900
p2.8xlarge	0.70	0.04	7.200
p2.16xlarge	0.85	0.04	14.400
g4dn.xlarge	0.47	0.02	0.526
g4dn.2xlarge	0.48	0.02	0.752
g4dn.4xlarge	0.49	0.03	1.204
g4dn.12xlarge	0.46	0.03	3.912
g4dn.16xlarge	0.47	0.03	4.352
g4dn.metal	0.47	0.03	7.824
g3.4xlarge	0.71	0.04	1.140
g3.8xlarge	0.64	0.03	2.280
g3.16xlarge	0.65	0.03	4.560
g3s.xlarge	0.50	0.03	0.750

fetches the most recent image from the AWS S3 bucket every 0.1 s. The platform's user interface comprises front-end and back-end. The front-end, directly shown to the user, is developed using TypeScript language with frameworks, such as Next.js and React.js, and the ant.design style library. The back-end, responsible for server processes, uses TypeScript with the Express.js framework, Object-Relational Mapping (ORM) with TypeORM, PostgreSQL DB, JsonWebToken for authentication, and the representational state transfer approach for the application programming interface.

For real-time monitoring, the time requirement of each process in the back-end is specifically validated and predicted. The video captured on the site needs to be downloaded in real time by AWS EC2 instances, the results of the multi-vision monitoring must be uploaded to an AWS S3 bucket, and then the uploaded images must be retrieved on the web-based platform. Initially, the authors estimated the time required to fetch a video captured on-site. An image with a 1920 × 1080 resolution consists of a total of 2073600 pixels, with each pixel containing information for red, green, and blue. Each color can have a value ranging from 0 to 255, requiring 1 byte of storage space per color. Therefore, an image of 1920 × 1080 resolution requires a storage size of 6220800 bytes, equivalent to 5.93 MB. Hence, for example, to analyze one video at three frames per second (fps), the on-site camera must upload data at a rate of 17.80 MB per second. Considering three video stream data, an upload speed of 53.40 MB per second is required. However, as of 2023, South Korea's 5G internet upload speed, despite being one of the world's internet powerhouses, is only 42.0 MB/s, equating to an upload capacity of about 5.25 MB/s [56]. Therefore, it is practically impossible to fetch high-definition videos captured by CCTV cameras installed at construction sites in real-time using wireless internet (e.g., 5G internet), and the research only considered wired internet. While wired internet services range from 500 MB/s to 10 GB/s, the authors assumed a 500 MB/s service for a cost-effective solution. The upload speed of a 500 MB/s wired internet service equates to 62.50 MB/s. This speed is sufficient to receive three fps videos in real time (53.40 MB/s < 62.50 MB/s), but not enough to receive four videos in real time.

4 Experimental results and discussion

To validate the proposed framework, the experiments were conducted on Sejong-Anseong highway construction site—Section 4 in South Korea on April 12, 2023. Three different videos recorded soil unloading zone (view 1), construction resources' traveling zone (view 2), and soil loading zone (view 3). For view 1, the camera recorded the unloading zone where dump trucks, dozers, and rollers were operating. Regarding view 2, the camera recorded the traveling zone, a connection zone between view 1 and view 3. View 3 was a video stream that an excavator and dump trucks were loading the soil. A total of 97080 image frames (view 1: 34520 images, view 2: 32720 images, view 3: 29840 images) were collected with 1920 × 1080, 1440 × 1080, and 1280 × 720 resolutions and three fps from the construction site. Since each zone has specific areas where tasks are primarily performed, cameras were strategically installed to collect video footage while ensuring that occlusion between resources within these work areas was minimized. The baseline DB and the site-optimized DB contained dump trucks, excavators, dozers, and rollers' information (e.g., visual characteristics). For the continuous multi-vision monitoring in real time, the training image DB development and model training were conducted on the AWS EC2 instances-g4dn.xlarge with an additional 500 GB (Ubuntu 20.04 LTS) storage for each view.

4.1 Object detection using real-time model updating

Object detection using real-time model updating consists of online-active learning and site-optimized DB. Before developing the site-optimized DB, a baseline DB is required. The baseline DB was built by web crawling and virtual reality, and 3424 images of baseline DB were pre-developed. Furthermore, the site-optimized DB of 3424 images was developed in 1100 s (Fig. 7). Considering that labeling images typically takes more than 10 s per image [32,48], the experiment demonstrated a time reduction of 96.79%, proving that the site-optimized DB can be used for real-time monitoring model updating.

When real-time model updating with the site-optimized DB was applied to each view, it achieved an average online-active learning operating frequency of 1114.36 s and an average macro F1-score of 87.30%. The longer frequency means the model's high performance is maintained for a longer time. The frequency and macro F1-score for each view was: view 1 1643.81 s and 84.81%, view 2 894.31 s and 86.84%, and view 3 804.95 s and 90.25%. Besides, when the real-time model updating used only the target site image (without site-optimized DB), it achieved an average online-active learning operating frequency of 37.29 s and an average macro F1-score of 46.00% (Table 3). The examples of object detection using real-time model updating results are Fig. 8. These results demonstrate that the proposed approach can be applied to various sites and resources, and the site-optimized DB can maintain the model's high performance for a longer duration.

4.2 Activity recognition

Activity recognition was carried out through the process of individual action recognition and interaction analysis. When the activity recognition model was applied to each view, it achieved an average accuracy of 86.20%. Specifically, the accuracy for each view was: view 1 81.47%, view 2 96.84%, and view 3 80.28% (Table 4, Figs. 9 and 10). In view 1, the altitude difference between the working area and the camera installation spot was not drastically different. This allowed for practical centroid movement analysis when construction resources moved perpendicularly to the camera's direction. However, it was hard to identify centroid movement when the movement was parallel to the camera's direction. This made distinguishing between 'move' and 'stop' less accurate. Consequently, there was more confusion between statuses that could be discerned from 'move' and 'stop' (e.g., a dump truck's travel versus idle) than between other activities. Additionally, in view 3, the construction resources were captured from a distance, making it detect and track smaller than the construction resources in view 1 and 2. This smaller bounding box led to errors when analyzing the interaction between different resources. This confused the excavator's activity status (i.e., load and soil-manage).

4.3 Multi-vision earthwork productivity monitoring

Multi-vision earthwork productivity monitoring was conducted by integrating the three different productivity monitoring results. The proposed multi-vision productivity monitoring result is as Table 5. In the case of view

Fig. 7 Example of the site-optimized DB generated by cross-oversampling the baseline DB and the target site image.

Table 3 Operating cycles of the model updating and macro F1-scores of the object detection models with and without site-optimized DB

Real-time model updating	View 1		View 2		View 3	
	Frequency (s)	Macro F1-score (%)	Frequency (s)	Macro F1-score (%)	Frequency (s)	Macro F1-score (%)
Without site-optimized DB	47.35	41.41	32.23	31.98	32.29	64.62
With site-optimized DB	1643.81	84.81	894.31	86.84	804.95	90.25

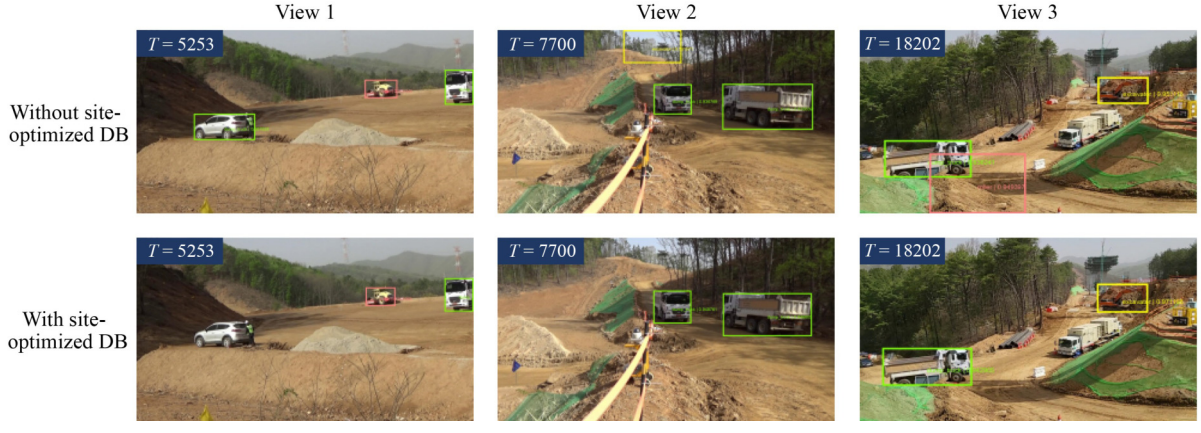

Fig. 8 Example of object detection using real-time monitoring model updating.

1, out of 29 dump trucks, six dump trucks were evaluated as outliers (travel: #3, #29, idle: #17, #26, unload: #24, #27). For instance, dump trucks #3 and #29 were analyzed to had taken longer travel time than other dump trucks (Fig. 11). Next, out of 71 dump trucks, three dump trucks were evaluated as outliers on view 2 (travel: #56, idle: #8, #52). To confirm the analysis result, the authors manually reviewed the video. The authors observed that dump truck #8 idled because it had to wait for an oncoming dump truck from the opposite direction, and it resulted in an extended idle time. Lastly, in the case of view 3, out of 30 dump trucks, six dump trucks were identified as outliers (travel: #9, idle: #18, #27, #29, #30, load: #13). Based on the authors' manual review, for example, dump truck #13 moved near the excavator to receive soil, but the excavator had not finished the soil management, causing the dump truck to remain idle for an extended duration. The dump truck was, however, too close to the excavator; it was recognized as being in a 'load' state (with the excavator in a 'semi-stop' state and the dump truck in a 'stop' state, and the interaction between the dump truck and the excavator recognized as yes), analyzed as in a longer load time (Fig. 12). The total earthwork volume is presented in Table 6. Specifically, it determined the number of dump trucks that performed loading operations, and through this, the authors could derive the total earthwork volume.

Table 4 Accuracy of the construction resource activity recognition

Resource type	View 1 (%)	View 2 (%)	View 3 (%)
Dump truck	77.32	96.84	80.68
Excavator	–	–	79.88
Dozer	71.76	–	–
Roller	95.33	–	–

For the multi-vision earthwork productivity monitoring, the productivity monitoring results of three viewpoints should be integrated to re-identify the IDs of the tracked

Fig. 9 Example of the activity recognition result.

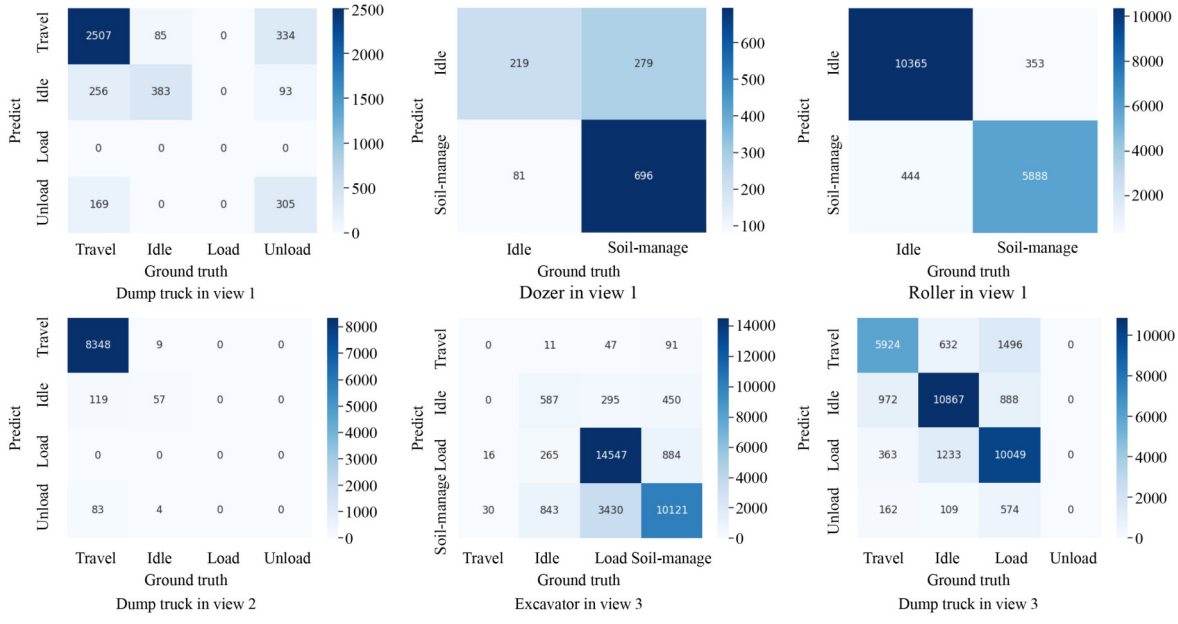

Fig. 10 Activity recognition confusion matrix.

Table 5 Productivity analysis result of dump trucks on each view

Item	ID	Appeared time	Disappeared time	Existed time	Travel time	Idle time	Unload time	Productivity ratio	Travel time *Z*-score	Idle time *Z*-score	Unload time *Z*-score
View 1	1	13.67	30.67	17.00	13.67	3.33	0.00	0.80	−0.69	−0.72	−0.75
	2	304.00	463.67	159.67	48.67	56.33	54.67	0.65	0.79	1.83	0.98
	3	661.00	824.00	163.00	78.67	23.33	61.00	0.86	2.06	0.24	1.18
	4	884.00	997.00	113.00	75.00	27.33	10.67	0.76	1.90	0.44	−0.41
	5	1264.00	1392.33	128.33	73.67	19.33	35.33	0.85	1.85	0.05	0.37
	6	2385.33	2398.00	12.67	9.00	3.33	0.33	0.74	−0.88	−0.72	−0.74
	7	2447.00	2461.33	14.33	11.00	3.33	0.00	0.77	−0.8	−0.72	−0.75
	8	3163.67	3174.67	11.00	6.33	3.67	1.00	0.67	−1.00	−0.70	−0.71
	9	3310.00	3323.33	13.33	9.33	3.33	0.67	0.75	−0.87	−0.72	−0.73
	10	3651.33	3670.33	19.00	13.33	5.00	0.67	0.74	−0.70	−0.64	−0.73
	11	3944.67	3959.67	15.00	11.00	3.67	0.33	0.76	−0.80	−0.70	−0.74
	12	4091.00	4101.67	10.67	6.33	4.00	0.33	0.62	−1.00	−0.69	−0.74
	13	4632.33	4648.67	16.33	11.33	4.33	0.67	0.73	−0.78	−0.67	−0.73
	14	4704.33	4736.67	32.33	18.00	8.33	6.00	0.74	−0.50	−0.48	−0.56
	15	4797.67	4808.67	11.00	7.33	3.33	0.33	0.70	−0.95	−0.72	−0.74
	16	4960.67	4976.67	16.00	8.67	7.00	0.33	0.56	−0.90	−0.54	−0.74
	17	5399.33	5525.33	126.00	44.33	63.67	18.00	0.49	0.61	2.18	−0.18
	18	6403.00	6449.33	46.33	16.33	27.33	2.67	0.41	−0.57	0.44	−0.66
	19	6524.00	6540.33	16.33	13.00	3.33	0.00	0.80	−0.71	−0.72	−0.75
	20	6783.00	6800.67	17.67	13.00	4.33	0.33	0.75	−0.71	−0.67	−0.74
	21	7115.33	7131.67	16.33	12.33	3.33	0.67	0.80	−0.74	−0.72	−0.73
	22	7244.33	7374.00	129.67	27.33	21.67	80.67	0.83	−0.11	0.16	1.80
	23	7510.33	7627.00	116.67	39.00	14.67	63.00	0.87	0.38	−0.17	1.24
	24	7793.00	7946.00	153.00	38.33	16.00	98.67	0.90	0.35	−0.11	2.37
	25	8001.00	8144.33	143.33	49.00	30.67	63.67	0.79	0.80	0.60	1.26
	26	8259.00	8450.33	191.33	45.67	90.33	55.33	0.53	0.66	3.47	1.00
	27	8630.33	8781.67	151.33	41.33	15.67	94.33	0.90	0.48	−0.13	2.23
	28	9019.00	9110.67	91.67	47.67	20.33	23.67	0.78	0.75	0.10	0.00
	29	9358.67	9490.67	132.00	79.33	39.67	13.00	0.70	2.08	1.03	−0.34
View 2	1	195.67	231.67	36.00	32.67	3.33	–	0.91	−0.02	−0.34	–
	2	283.00	323.33	40.33	36.67	3.67	–	0.91	0.41	−0.24	–
	3	400.67	442.67	42.00	35.33	6.67	–	0.84	0.27	0.62	–
	4	552.33	600.33	48.00	44.00	4.00	–	0.92	1.20	−0.15	–
	5	785.67	834.67	49.00	45.33	3.67	–	0.93	1.34	−0.24	–
	6	799.33	831.00	31.67	25.00	6.67	–	0.79	−0.84	0.62	–
	7	879.67	911.67	32.00	28.67	3.33	–	0.90	−0.45	−0.34	–
	8	1025.33	1099.33	74.00	46.33	27.67	–	0.76	1.45	6.64	–
	9	1050.00	1087.67	37.67	34.00	3.67	–	0.90	0.12	−0.24	–
	10	1182.33	1229.67	47.33	43.67	3.67	–	0.92	1.16	−0.24	–
	11	1232.00	1263.00	31.00	27.67	3.33	–	0.89	−0.56	−0.34	–
	12	1331.67	1371.67	40.00	36.67	3.33	–	0.92	0.41	−0.34	–
	13	1507.33	1542.67	35.33	31.67	3.67	–	0.90	−0.13	−0.24	–
	14	1560.00	1617.00	57.00	51.00	6.00	–	0.89	1.95	0.43	–
	15	1697.00	1727.00	30.00	26.67	3.33	–	0.89	−0.66	−0.34	–

(*Continued*)

Item	ID	Appeared time	Disappeared time	Existed time	Travel time	Idle time	Unload time	Productivity ratio	Travel time *Z*-score	Idle time *Z*-score	Unload time *Z*-score
View 2	16	1798.67	1834.00	35.33	27.33	8.00	–	0.77	−0.59	1.00	–
	17	1858.00	1885.67	27.67	23.33	4.33	–	0.84	−1.02	−0.05	–
	18	2076.00	2106.67	30.67	27.33	3.33	–	0.89	−0.59	−0.34	–
	19	2265.33	2275.33	10.00	6.67	3.33	–	0.67	−2.81	−0.34	–
	20	2384.33	2411.00	26.67	22.33	4.33	–	0.84	−1.13	−0.05	–
	21	2564.00	2593.67	29.67	26.33	3.33	–	0.89	−0.70	−0.34	–
	22	2580.33	2609.67	29.33	22.67	6.67	–	0.77	−1.09	0.62	–
	23	2899.33	2938.33	39.00	35.67	3.33	–	0.91	0.30	−0.34	–
	24	2925.67	2961.00	35.33	32.00	3.33	–	0.91	−0.09	−0.34	–
	25	3059.33	3078.33	19.00	15.00	4.00	–	0.79	−1.91	−0.15	–
	26	3082.33	3109.00	26.67	23.33	3.33	–	0.87	−1.02	−0.34	–
	27	3160.67	3209.00	48.33	42.00	6.33	–	0.87	0.98	0.52	–
	28	3166.67	3201.67	35.00	28.33	6.67	–	0.81	−0.48	0.62	–
	29	3322.33	3364.67	42.33	38.67	3.67	–	0.91	0.62	−0.24	–
	30	3368.67	3402.33	33.67	30.33	3.33	–	0.90	−0.27	−0.34	–
	31	3503.00	3525.33	22.33	19.00	3.33	–	0.85	−1.49	−0.34	–
	32	3815.00	3859.00	44.00	40.33	3.67	–	0.92	0.80	−0.24	–
	33	3983.33	4030.00	46.67	43.00	3.67	–	0.92	1.09	−0.24	–
	34	3997.67	4008.33	10.67	7.33	3.33	–	0.69	−2.74	−0.34	–
	35	4110.67	4141.33	30.67	27.33	3.33	–	0.89	−0.59	−0.34	–
	36	4187.00	4219.00	32.00	28.67	3.33	–	0.90	−0.45	−0.34	–
	37	4221.00	4251.33	30.33	27.00	3.33	–	0.89	−0.63	−0.34	–
	38	4390.00	4419.00	29.00	25.67	3.33	–	0.89	−0.77	−0.34	–
	39	4617.00	4664.33	47.33	43.33	4.00	–	0.92	1.12	−0.15	–
	40	4640.33	4683.33	43.00	39.67	3.33	–	0.92	0.73	−0.34	–
	41	4861.67	4896.67	35.00	31.67	3.33	–	0.90	−0.13	−0.34	–
	42	4938.67	4978.67	40.00	36.67	3.33	–	0.92	0.41	−0.34	–
	43	5608.33	5649.33	41.00	37.67	3.33	–	0.92	0.52	−0.34	–
	44	5767.33	5803.67	36.33	32.67	3.67	–	0.90	−0.02	−0.24	–
	45	5856.00	5901.33	45.33	41.33	4.00	–	0.91	0.91	−0.15	–
	46	5944.67	5981.00	36.33	33.00	3.33	–	0.91	0.02	−0.34	–
	47	5981.67	6032.00	50.33	46.33	4.00	–	0.92	1.45	−0.15	–
	48	6213.00	6248.00	35.00	31.00	4.00	–	0.89	−0.20	−0.15	–
	49	6313.67	6351.33	37.67	34.33	3.33	–	0.91	0.16	−0.34	–
	50	6420.00	6463.00	43.00	39.67	3.33	–	0.92	0.73	−0.34	–
	51	6531.67	6558.33	26.67	23.33	3.33	–	0.87	−1.02	−0.34	–
	52	6539.33	6593.00	53.67	34.33	19.33	–	0.64	0.16	4.25	–
	53	6560.00	6582.33	22.33	19.00	3.33	–	0.85	−1.49	−0.34	–
	54	6584.00	6604.33	20.33	17.00	3.33	–	0.84	−1.70	−0.34	–
	55	6784.33	6820.00	35.67	29.00	6.67	–	0.81	−0.41	0.62	–
	56	6797.67	6857.67	60.00	56.67	3.33	–	0.94	2.56	−0.34	–
	57	6859.33	6893.00	33.67	30.33	3.33	–	0.90	−0.27	−0.34	–
	58	7033.00	7066.33	33.33	30.00	3.33	–	0.90	−0.31	−0.34	–
	59	7096.00	7143.33	47.33	43.33	4.00	–	0.92	1.12	−0.15	–

(*Continued*)

Item	ID	Appeared time	Disappeared time	Existed time	Travel time	Idle time	Unload time	Productivity ratio	Travel time *Z*-score	Idle time *Z*-score	Unload time *Z*-score
View 2	60	7313.33	7354.33	41.00	37.67	3.33	–	0.92	0.52	−0.34	–
	61	7355.67	7400.33	44.67	40.67	4.00	–	0.91	0.84	−0.15	–
	62	7545.00	7583.33	38.33	31.67	6.67	–	0.83	−0.13	0.62	–
	63	7562.00	7609.33	47.33	43.00	4.33	–	0.91	1.09	−0.05	–
	64	7860.33	7892.33	32.00	28.67	3.33	–	0.90	−0.45	−0.34	–
	65	7313.33	7980.00	47.00	43.67	3.33	–	0.93	1.16	−0.34	–
	66	8194.00	8236.00	42.00	38.00	4.00	–	0.90	0.55	−0.15	–
	67	8332.67	8362.33	29.67	26.33	3.33	–	0.89	−0.70	−0.34	–
	68	8511.67	8548.33	36.67	33.00	3.67	–	0.90	0.02	−0.24	–
	69	8658.67	8705.67	47.00	43.67	3.33	–	0.93	1.16	−0.34	–
	70	8902.67	8938.00	35.33	31.67	3.67	–	0.90	−0.13	−0.24	–
	71	8985.67	9027.67	42.00	38.67	3.33	–	0.92	0.62	−0.34	–
View 3	1	109.67	308.00	198.33	51.33	26.67	121.33	0.87	−1.87	−0.61	−1.11
	2	355.33	555.33	200.00	90.33	0.00	168.00	1.00	0.69	−0.82	−0.30
	3	577.33	855.33	278.00	53.67	0.00	242.67	1.00	−1.72	−0.82	1.00
	4	823.67	1088.33	264.67	80.00	30.00	154.67	0.89	0.01	−0.59	−0.53
	5	1134.33	1315.00	180.67	75.67	0.00	138.67	1.00	−0.27	−0.82	−0.81
	6	1253.33	1492.67	239.33	50.00	52.33	181.33	0.82	−1.96	−0.41	−0.07
	7	1596.00	1809.00	213.00	90.33	0.00	193.33	1.00	0.69	−0.82	0.14
	8	1864.67	2145.33	280.67	78.00	80.00	142.00	0.73	−0.12	−0.19	−0.75
	9	1997.00	2368.33	371.33	114.33	138.33	159.00	0.66	2.27	0.27	−0.45
	10	2304.00	2648.33	344.33	91.00	46.00	230.33	0.87	0.74	−0.46	0.78
	11	2698.33	2942.67	244.33	75.00	18.00	177.67	0.93	−0.32	−0.68	−0.13
	12	2882.00	3136.67	254.67	59.00	51.67	165.67	0.81	−1.37	−0.42	−0.34
	13	3126.33	3566.00	439.67	94.00	0.00	354.67	1.00	0.93	−0.82	2.94
	14	3439.00	3785.33	346.33	84.67	145.00	148.33	0.62	0.32	0.32	−0.64
	15	3726.00	3996.00	270.00	77.67	50.67	169.67	0.83	−0.14	−0.42	−0.27
	16	4145.33	4365.00	219.67	74.00	0.00	160.67	1.00	−0.38	−0.82	−0.42
	17	4944.33	5394.00	449.67	70.33	245.00	178.33	0.50	−0.62	1.11	−0.12
	18	4983.00	5570.00	587.00	66.00	400.67	150.33	0.35	−0.91	2.34	−0.60
	19	5483.33	5741.00	257.67	97.67	166.67	272.33	0.69	1.18	0.49	1.51
	20	6064.33	6244.33	180.00	69.00	0.00	163.67	1.00	−0.71	−0.82	−0.37
	21	6119.67	6508.33	388.67	87.67	170.00	154.33	0.59	0.52	0.52	−0.53
	22	6394.67	6700.33	305.67	91.33	94.00	221.00	0.77	0.76	−0.08	0.62
	23	6827.00	7026.00	199.00	81.33	0.00	185.33	1.00	0.10	−0.82	0.00
	24	6945.00	7256.33	311.33	70.67	68.33	225.33	0.81	−0.60	−0.28	0.70
	25	7218.33	7460.00	241.67	100.00	17.67	169.67	0.94	1.33	−0.68	−0.27
	26	7632.00	7673.00	41.00	68.00	0.00	166.00	1.00	−0.78	−0.82	−0.33
	27	7702.00	8281.33	579.33	74.67	368.33	210.33	0.44	−0.34	2.08	0.44
	28	8105.67	8556.00	450.33	103.00	226.67	132.00	0.51	1.53	0.96	−0.92
	29	8432.00	8906.33	474.33	91.33	386.00	77.00	0.30	0.76	2.22	−1.88
	30	8721.00	9278.00	557.00	84.00	351.67	340.00	0.55	0.28	1.95	2.69

resources. A total of 88560 images were used for the experiments, with 29520 images from each view, covering the same footage from 08:42 to 11:26 AM under changing sunrise conditions. Table 7 presents the re-

Fig. 11 Example of relatively long travel time (dump truck #3 in view 1).

Fig. 12 Dump truck idles near the soil managing excavator (dump truck #13 in view 3).

Table 6 Dump truck earthwork volume analyzed result

Contents for earthwork volume analysis	Value
Number of dump trucks	30
Operated time	9508 s
Dump truck capacity	15 t
Earthwork volume	450 t

Table 7 Dump truck re-identification result

Zone change	Number of re-identified IDs	Number of total IDs	Re-identification accuracy (%)
Zone 1 → Zone 2	23	24	95.83
Zone 2 → Zone 1	23	27	85.19
Zone 2 → Zone 3	27	27	100
Zone 3 → Zone 2	28	28	100

identification results for multi-vision monitoring. Among four types of resources, only dump trucks were observed to move between different views. Consequently, the analysis focused on the re-identification of dump trucks across four scenarios: moving from zone 1 to 2, 2 to 1, 2 to 3, and 3 to 2. The average accuracy of the re-identification process was 95.26%, with accuracies of 95.83% for zone 1 to 2, 85.19% for zone 2 to 1, and 100% for both zone 2 to 3 and zone 3 to 2. Occlusion between resources occurred continuously in zone 2. Despite being a transition zone that connects two different zones and having frequent movement of various resources, a high re-identification accuracy was observed between zone 2 and zone 3. However, the accuracy related to zone 1 was relatively low and it can be attributed to the periodic changes in the location where the dump truck unloads, which make it difficult to record all relevant scenes in the video.

To integrate the IDs, the minimum and maximum duration of traveling from view 1 to 2 was set as five seconds and 100 s, and view 2 to 3 was set as 160 and 330 s by considering the conditions of the target construction site (e.g., distance between the zones). The IDs were successfully integrated, and it was confirmed that three dump trucks were successfully and continuously tracked (Table 8). They were named as dump truck A, B, and C.

Based on the ID integration result, multi-vision earthwork productivity monitoring was conducted (Table 9).

Table 8 Details of the integrated dump truck IDs

Resource ID	View	Dump truck IDs in each view																	
Dump truck A	view 1	3	6	8	11	15	19	22	25	28									
	view 2	2	5	8	14	16	20	22	24	30	34	37	44	46	53	55	60	62	67	69
	view 3	3	6	9	12	15	18	21	24	27									
Dump truck B	view 1	4	7	9	12	17	20	23	26										
	view 2	1	3	9	11	15	17	21	29	31	39	42	47	48	58	63	64	66	71
	view 3	1	4	7	10	13	16	19	22	25	29								
Dump truck C	view 1	5	10	13	18	21	24	27											
	view 2	4	6	12	13	18	19	23	26	32	35	43	45	49	52	59	61	65	68
	view 3	2	5	8	11	14	17	20	23	26	30								

Table 9 Multi-vision earthwork productivity monitoring result

Dump truck	View 1			View 2		View 3			Productivity ratio
	Travel	Idle	Unload	Travel	Idle	Travel	Idle	Load	
A	249.33	113.33	231.00	601.33	106.99	653.68	1300.33	1658.66	0.69
B	243.66	210.99	148.33	599.34	68.64	860.99	767.01	1955.00	0.78
C	206.65	90.99	233.01	579.00	79.65	776.33	839.67	1808.00	0.78

To compare the performance of three dump trucks, the productivity ratio (productive work time ratio) was calculated. As a result, dump trucks B and C showed the highest productivity ratio, while dump truck A had the lowest. Notably, dump truck A had a longer idle time that was almost twice as long as that of other dump trucks. It confirmed that this outcome was due to dump truck A continuously idling before loading in view 3.

4.4 Site monitoring via the web-based platform

The multi-vision monitoring results were uploaded and displayed in real-time on a web-based platform. Using the YOLOv5-based real-time monitoring model updating, images with a resolution of 1920 × 1080 were analyzed on average within 0.02 s. To exclude any delay during analysis, the authors assumed the required time to analyze the image as maximum 0.04 s. Based on the detection results, tracking was utilized to assign IDs, and through both individual action and interaction analysis, the activity recognition process was completed within a maximum of 0.01 s per frame. Moreover, the time taken to integrate the vision-based monitoring results for each video was, on average, 0.005 s per frame (Fig. 13).

All subsequent processes occurred within the AWS environment. It took approximately 0.1 s to upload the multi-vision monitoring result images to the S3 bucket and about 0.1 s to retrieve them in AWS Lightsail. In summary, the total time required to analyze the image frames and display the real-time monitoring results on the platform was, on average, 0.255 s. This is shorter than 0.33 s required to download an image frame from 3 fps real-time video, verifying that the analysis can be done in real time. The users can check the displayed productivity monitoring results by resources as illustrated in Fig. 14.

In summary, the object detection model achieved a real-time model updating frequency of 1114.36 s and a macro F1-score of 87.30%. Based on the object detection results; despite analyzing the activities of four types of construction resources in three different viewpoints, it demonstrated a sufficient performance on activity recognition with an accuracy of 86.20%. Furthermore, multi-vision earthwork productivity monitoring was successfully carried out by re-identification of IDs with an accuracy of 95.26%. Lastly, the results of multi-vision productivity monitoring were successfully displayed through the web-based platform in real time. Although this study focused on three site videos and four construction resource types for validation, the framework can be applied to other sites directly by re-building site-optimized DB and changing minor parameters such as modifying the site condition (e.g., site layout).

Fig. 13 Example of displaying real-time monitoring results on the web-based platform.

5 Conclusions

In this study, the authors proposed a sustainable framework for multi-vision earthwork productivity monitoring. The proposed framework involves four processes: 1) site-optimized DB development; 2) real-time monitoring model updating; 3) multi-vision

Fig. 14 Web page to display the productivity monitoring results by resources.

productivity monitoring; and 4) web-based monitoring platform for Internet-connected devices. To validate the proposed framework, the authors conducted the experiments. Using the site-optimized DB allowed to increase the model update frequency by 29.88 times and improve the macro F1-score by 41.30% for detection. Activity recognition for four types of resources achieved a macro F1-score of 86.20%, and successful vision-based productivity monitoring was carried out on three videos, which were then integrated to realize multi-vision earthwork productivity. Furthermore, these monitoring results were successfully displayed on the web-based platform, enabling multi-vision site monitoring without spatiotemporal constraints.

Given the benefits of the proposed framework, this study makes the following contributions. First, this study proposed a generalized framework to apply vision-based monitoring to any other construction site without relying on camera viewpoints and target resources. The proposed framework was able to customize the monitoring model regardless of the site conditions (e.g., characteristic of the image background). Second, it became possible to update the vision-based monitoring models in real time. This can enable the rapid application of computer vision techniques to construction sites. Third, the authors integrated vision-based monitoring results from multiple videos by considering the conditions of the construction site. Fourth, real-time analysis of huge and dynamic construction sites became possible, and the results can be monitored on various devices. Lastly, this research can support other research that requires long-term monitoring (e.g., progress monitoring) by real-time monitoring model updating.

Although interesting findings were observed in this study, there are still limitations to be addressed. For instance, we used videos captured outdoors during daylight hours, ensuring adequate brightness for vision-based monitoring. As a result, we did not specifically consider variations in lighting conditions. However, it is important to note that most construction activities are conducted during daylight hours for safety and operational efficiency, minimizing the impact of extreme lighting changes. According to Occupational Safety and Health Administration Standard 1926.56(a) [57], earthwork sites must maintain an illumination level of at least three foot-candles to ensure safe working conditions. This reinforces the assumption that lighting conditions in typical construction environments are generally adequate for vision-based monitoring. Nonetheless, certain construction environments may require greater robustness against lighting variations. Future research can explore adaptive techniques to improve system performance under diverse lighting conditions. Additionally, this study specifically focused on earthwork, where large construction equipment plays a dominant role. Earthwork sites typically have relatively structured movement patterns, making object re-identification more feasible and effective. However, generalizing the proposed framework to different site conditions, such as small size earthwork sites in a city cluttered with the increased number of workers or earthwork sites in highway work zones causing more complex object interactions, is also important. Future research could explore methods to enhance object re-identification under such conditions.

Based on the findings, additional research opportunities are presented. First, with maintaining the high

performance of the monitoring model, it is no longer about applying state-of-the-art vision technologies to construction sites in the short-term. Instead, applying them consistently over the longer term can maximize the potential for utilizing computer vision technologies in construction sites. Second, by collecting the location or activity information of each resource continuously, it is possible to develop a digital twin that represents the construction site in a 3D virtual space in real time. Continuously implementing a construction site into a digital twin in real time allows for the simulation of all scenarios in advance. Furthermore, construction productivity can be enhanced for a wider range of heavy operations, and their safety can be also investigated. The authors believe that with more such achievements, real-time multi-vision monitoring on construction sites can lead to future innovative and intelligent construction sites.

Acknowledgements This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (Nos. RS-2023-00241758, 2021R1A2C2003696, and RS-2024-00334513).

Funding note Open access funding provided by Seoul National University.

Competing interests The authors declare that they have no competing interests.

References

1. Delgado J M D, Oyedele L, Ajayi A, Akanbi L, Akinade O, Bilal M, Owolabi H. Robotics and automated systems in construction: Understanding industry-specific challenges for adoption. Journal of Building Engineering, 2019, 26: 100868
2. Pan W, Chen L, Zhan W. PESTEL analysis of construction productivity enhancement strategies: A case study of three economies. Journal of Management Engineering, 2019, 35(1): 05018013
3. Ghodrati N, Yiu T W, Wilkinson S, Shahbazpour M. Role of management strategies in improving labor productivity in general construction projects in New Zealand: Managerial perspective. Journal of Management Engineering, 2018, 34(6): 04018035
4. Bankvall L, Bygballe L E, Dubois A, Jahre M. Interdependence in supply chains and projects in construction. Supply Chain Management, 2010, 15(5): 385–393
5. Barbosa F, Woetzel J, Mischke J, Ribeirinho M J, Sridhar M, Parsons M, Bertram N, Brown S. Reinventing Construction: A Route to Higher Productivity. Brussels: McKinsey & Company, 2017
6. Kang S H, Seo J W, Baik K G. 3D-GIS based earthwork planning system for productivity improvement. In: Proceedings of the Construction Research Congress 2009: Building a Sustainable Future. Seattle, WA: ASCE, 2012,151–160
7. Wong E, Swei O. New construction cost indices to improve highway management. Journal of Management Engineering, 2021, 37(4): 04021030
8. Kisi K P, Mani N, Rojas E M, Foster E T. Optimal productivity in labor-intensive construction operations: Pilot study. Journal of Construction Engineering and Management, 2017, 143(3): 04016107
9. Chen C, Zhu Z, Hammad A. Automated excavators activity recognition and productivity analysis from construction site surveillance videos. Automation in Construction, 2020, 110: 103045
10. Wu H, Zhong B, Li H, Guo J, Wang Y. On-site construction quality inspection using blockchain and smart contracts. Journal of Management Engineering, 2021, 37(6): 04021065
11. Wang S H. Engineering productivity and unit price assessment model. Journal of Management Engineering, 2022, 38(1): 04021076
12. Pradhananga N, Teizer J. Automatic spatio-temporal analysis of construction site equipment operations using GPS data. Automation in Construction, 2013, 29: 107–122
13. Lu W, Huang G Q, Li H. Scenarios for applying RFID technology in construction project management. Automation in Construction, 2011, 20(2): 101–106
14. Zhang C, Shen W, Ye Z. Technical feasibility analysis on applying ultra-wide band technology in construction progress monitoring. International Journal of Construction Management, 2022, 22(15): 2951–2965
15. Gu Y, Ai Q, Xu Z, Yao L, Wang H, Huang X, Yuan Y. Cost-effective image recognition of water leakage in metro tunnels using self-supervised learning. Automation in Construction, 2024, 167: 105678
16. Ai Q, Yuan Y, Bi X. Acquiring sectional profile of metro tunnels using charge-coupled device cameras. Structure and Infrastructure Engineering, 2016, 12(9): 1065–1075
17. Ai Q, Yuan Y. Rapid acquisition and identification of structural defects of metro tunnel. Sensors, 2019, 19(19): 4278
18. Kim J, Hwang J, Jeong I, Chi S, Seo J O, Kim J. Generalized vision-based framework for construction productivity analysis using a standard classification system. Automation in Construction, 2024, 165: 105504
19. Kim H, Kim H, Hong Y W, Byun H. Detecting construction equipment using a region-based fully convolutional network and transfer learning. Journal of Computing in Civil Engineering, 2018,

32(2): 04017082
20. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149
21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016. New York, NY: IEEE, 2016, 770–778
22. Li Y, Lu Y, Chen J. A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector. Automation in Construction, 2021, 124: 103602
23. Shin Y, Choi Y, Won J, Hong T, Koo C. A new benchmark model for the automated detection and classification of a wide range of heavy construction equipment. Journal of Management Engineering, 2024, 40(2): 04023069
24. Nath N D, Behzadan A H. Deep convolutional networks for construction object detection under different visual conditions. Frontiers in Built Environment, 2020, 6: 97
25. Zhu Z, Ren X, Chen Z. Integrated detection and tracking of workforce and equipment from construction jobsite videos. Automation in Construction, 2017, 81: 161–171
26. Xiao B, Zhu Z. Two-dimensional visual tracking in construction scenarios: A comparative study. Journal of Computing in Civil Engineering, 2018, 32(3): 04018006
27. Jeong I, Hwang J, Kim J, Chi S, Hwang B G, Kim J. Vision-based productivity monitoring of tower crane operations during curtain wall installation using a database-free approach. Journal of Computing in Civil Engineering, 2023, 37(4): 04023015
28. Xiao B, Kang S C. Vision-based method integrating deep learning detection for tracking multiple construction machines. Journal of Computing in Civil Engineering, 2021, 35(2): 04020071
29. Angah O, Chen A Y. Tracking multiple construction workers through deep learning and the gradient based method with re-matching based on multi-object tracking accuracy. Automation in Construction, 2020, 119: 103308
30. Yan X, Zhang H, Gao H. Mutually coupled detection and tracking of trucks for monitoring construction material arrival delays. Automation in Construction, 2022, 142: 104491
31. Luo H, Xiong C, Fang W, Love P E D, Zhang B, Ouyang X. Convolutional neural networks: Computer vision-based workforce activity assessment in construction. Automation in Construction, 2018, 94: 282–289
32. Kim J, Chi S. Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles. Automation in Construction, 2019, 104: 255–264
33. Zhang J, Zi L, Hou Y, Wang M, Jiang W, Deng D. A deep learning-based approach to enable action recognition for construction equipment. Advances in Civil Engineering, 2020, 2020(1): 8812928
34. Luo X, Li H, Cai D, Dai F, Seo H O, Lee S H. Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks. Journal of Computing in Civil Engineering, 2018, 32(3): 04018012
35. Roberts D, Golparvar-Fard M. End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level. Automation in Construction, 2019, 105: 102811
36. Soltani M M, Zhu Z, Hammad A. Framework for location data fusion and pose estimation of excavators using stereo vision. Journal of Computing in Civil Engineering, 2018, 32(6): 04018045
37. Cheng J C P, Wong P K Y, Luo H, Wang M, Leung P H. Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. Automation in Construction, 2022, 139: 104312
38. Zhang Q, Wang Z, Yang B, Lei K, Zhang B, Liu B. Reidentification-based automated matching for 3D localization of workers in construction sites. Journal of Computing in Civil Engineering, 2021, 35(6): 04021019
39. Wei R, Love P E D, Fang W, Luo H, Xu S. Recognizing people's identity in construction sites with computer vision: A spatial and temporal attention pooling network. Advanced Engineering Informatics, 2019, 42: 100981
40. Kim H, Bang S, Jeong H, Ham Y, Kim H. Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation. Automation in Construction, 2018, 92: 188–198
41. Kim H, Ham Y, Kim W, Park S, Kim H. Vision-based nonintrusive context documentation for earthmoving productivity simulation. Automation in Construction, 2019, 102: 135–147
42. Cheng M Y, Cao M T, Nuralim C K. Computer vision-based deep learning for supervising excavator operations and measuring real-time earthwork productivity. Journal of Supercomputing, 2023, 79(4): 4468–4492
43. Bügler M, Borrmann A, Ogunmakin G, Vela P A, Teizer J. Fusion of photogrammetry and video analysis for productivity assessment of earthwork processes. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(2): 107–123
44. Chen C, Xiao B, Zhang Y, Zhu Z. Automatic vision-based calculation of excavator earthmoving productivity using zero-shot learning activity recognition. Automation in Construction, 2023, 146: 104702
45. Hwang J, Kim J, Chi S, Seo J O. Development of training image database using web crawling for vision-based site monitoring. Automation in Construction, 2022, 135: 104141
46. Lee J G, Hwang J, Chi S, Seo J. Synthetic image dataset development for vision-based construction equipment detection. Journal of Computing in Civil Engineering, 2022, 36(5): 04022020
47. Hwang J, Kim J, Chi S. Site-optimized training image database development using web-crawled and synthetic images. Automation in Construction, 2023, 151: 104886
48. Kim J, Hwang J, Chi S, Seo J O. Towards database-free vision-based monitoring on construction sites: A deep active learning approach. Automation in Construction, 2020, 120: 103376
49. Meng D, Yang S, de Jesus A M P, Zhu S P. A novel Kriging-model-assisted reliability-based multidisciplinary design optimization strategy and its application in the offshore wind turbine tower. Renewable Energy, 2023, 203: 407–420
50. Meng D, Yang S, de Jesus A M P, Fazeres-Ferradosa T, Zhu S P. A novel hybrid adaptive Kriging and water cycle algorithm for reliability-based design and optimization strategy: Application in offshore wind turbine monopile. Computer Methods in Applied Mechanics and Engineering, 2023, 412: 116083

51. Meng D, Yang H, Yang S, Zhang Y, de Jesus A M P, Correia J, Fazeres-Ferradosa T, Macek W, Branco R, Zhu S P. Kriging-assisted hybrid reliability design and optimization of offshore wind turbine support structure based on a portfolio allocation strategy. Ocean Engineering, 2024, 295: 116842
52. Yang S, Meng D, Wang H, Yang C. A novel learning function for adaptive surrogate-model-based reliability evaluation. Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences, 2024, 382(2264): 20220395
53. Kim J, Chi S. Adaptive detector and tracker on construction sites using functional integration and online learning. Journal of Computing in Civil Engineering, 2017, 31(5): 04017026
54. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, NY: IEEE, 2016, 779–788
55. Park C, Chun H, Chi S. Multi-camera people counting using a queue-buffer algorithm for effective search and rescue in building disasters. KSCE Journal of Civil Engineering, 2024, 28(6): 2132–2146
56. Fogg I. Benchmarking the Global 5G Experience. Opensignal Limited 2023 June. 2023
57. Occupational Safety and Health Administration Standard 1926. Safety and Health Regulations for Construction. Washington, D.C.: OSHA, 1926