Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future

Jun Yang, Man-Woo Park, Patricio A. Vela, Mani Golparvar-Fard

Article history:
Received 24 September 2014
Received in revised form 12 December 2014
Accepted 22 January 2015
Available online 20 February 2015

Keywords:
Construction
Performance monitoring
Computer vision
Machine learning

ABSTRACT

Timely and accurate monitoring of onsite construction operations can bring an immediate awareness on project specific issues. It provides practitioners with the information they need to easily and quickly make project control decisions. Despite their importance, the current practices are still time-consuming, costly, and prone to errors. To facilitate the process of collecting and analyzing performance data, researchers have focused on devising methods that can semi-automatically or automatically assess ongoing operations both at project level and operation level. A major line of work has particularly focused on developing computer vision techniques that can leverage still images, time-lapse photos and video streams for documenting the work in progress. To this end, this paper extensively reviews these state-of-the-art vision-based construction performance monitoring methods. Based on the level of information perceived and the types of output, these methods are mainly divided into two categories (namely project level: visual monitoring of civil infrastructure or building elements vs. operation level: visual monitoring of construction equipment and workers). The underlying formulations and assumptions used in these methods are discussed in detail. Finally the gaps in knowledge that need to be addressed in future research are identified.

1. Introduction

Systematic monitoring of onsite construction operations can bring an immediate awareness on project specific issues [1–4]. It provides practitioners – owners, contractors, subcontractors, and tradesmen – with the information they need to easily and quickly make project control decisions [5,6]. Onsite project monitoring typically includes tracking project-level information, and also direct observations, surveys and interviews for monitoring operation-level information. While these methods are complementary, each of them is designed and engineered to measure a certain aspect of the construction production. The practices of implementing these methods also vary significantly across different construction firms.

Monitoring project-level information involves characterizing the extent to which construction plans are being followed. More specifically, it highlights when performance deviations happen and refers to tracking progress on construction of the civil infrastructure or building elements and monitoring the quality of work being executed. The outcome of these assessments includes unit rates which are then related to performance metrics such as Schedule Performance Index (SPI) in Earned Value Analysis or Plan Percent Complete (PPC) in Last Planner System. As such, these methods only reveal issues pertaining to the global outcome in production and do not involve measuring the performance of the utilized construction equipment and workers.

To measure construction production at the operational level, both onsite observations and survey or interview based methods are used [7–10]. These methods characterize the extent to which workers and equipment are being fully utilized and are operated under a safe work environment (a process formally known as activity monitoring). By tracking movements and activities of equipment
and craft workers and relating those to performance metrics such as productivity in form of crew-balance charts — together with surveys and interviews — operation-level monitoring highlights some of the root-causes of why performance deviations happen at the operation level.

Although a framework for monitoring both project-level and operation-level information offers a plausible solution to improve onsite operations, yet one of the major limitations is that implementing current methods is time-consuming, costly, and prone to errors. More specifically, the current practices in progress monitoring lack the human–machine interactivity. Currently most queries of the site situation are conducted by (1) traveling between the site and trailers to access paper-based drawings and specifications to generate an updated base for assessments or (2) at best manually accessing and searching through the plan through tablets which requires specific 3D rendered views of the plan to be manually generated beforehand for each inspection task [11–14]. This process is very time-consuming given thousands of elements and hundreds of construction activities on a site.

Also analyzing performance deviations between the construction plan and the as-built performance primarily relies on experience of the onsite inspectors which makes it subjective and often error prone (see Fig. 1). Furthermore, the need for additional information during a post-processing phase may require contacting field engineers offsite or the field inspectors to go back to the site and conduct additional data collections at a later stage in which the state of the actual progress might have already changed.

The situation is the same with the current practice of activity monitoring and analysis. Site activities involve numerous parties such as contractors, subcontractors, and trades, who may require equipment to complete their tasks. Managing these activities to achieve maximum operational efficiency from the required resources is difficult and as reported by the National Research Council [15] is not typically done well. The manual implementation of direct observations is also time-consuming, labor-intensive, and can be prone to errors. The significant amount of information which is required to be manually collected from various locations on a construction site may adversely affect the quality of the analysis, and according to the U.S. National Institute of Standards and Technology, minimizes the opportunity for continuous benchmarking and monitoring which is a necessary step for performance improvement [16]. In addition, these models can only provide post-process evaluations on construction activities. Such post-evaluation models are only applicable to the enterprises in which the manufacturing process is repeated routinely. In the case of one-time running state of construction projects, without automated and near real-time data collection, such techniques will have limited benefits [17–20].

For both project-level and also operation-level monitoring, collecting data and transforming it to information takes away from the more important tasks of identifying opportunities for performance improvements, reviewing alternatives, and conducting what-if analysis. Considering these challenges, the need for “automation” in all of these fronts becomes more clear and is in fact also highlighted by the U.S. National Academy of Engineering. During the past decade, advanced information and communication technologies (ICT) have been increasingly applied to construction projects to address current limitations and provide automatic data collection and analysis. As a type of easily captured and widely spread media, images and videos have become prevalent on common construction sites. Applying computer vision technology to analyze the recorded images and videos automatically has drawn much attention from both civil engineers and computer scientists these years. Collaborations between computer vision and civil engineering researchers and also several interdisciplinary efforts have enabled the measuring, detecting and tracking of civil infrastructure elements, equipment and workers, all of which play critical roles in construction performance monitoring applications, such as progress monitoring, quality control, operation analysis, safety monitoring and occupational health assessments [2,18].

Following is a comprehensive literature review on the state-of-the-art of vision-based construction performance monitoring, which will be structured based on the level of information perceived together with the corresponding outputs (namely project level: civil infrastructure or building elements vs. operation level: equipment and workers). Visual monitoring of civil infrastructure or building elements involves recognizing the state (both in geometry and appearance) of a civil infrastructure or a building element over time and compare the as-built state against the as-planned model and schedule. The input to this type of application is often still-images or time-lapse photos. Visual monitoring of equipment and workers requires detecting, tracking their location in 2D or 3D and analyzing their activities eventually, which is usually based on the onsite captured video streams. The various aspects for each of these groups of the literature are discussed in detail. In particular, based on the literature review, the paper presents open challenges in this field and suggests new directions for future research.

2. Visual monitoring of civil infrastructure or building elements

Today the availability of inexpensive point-and-shoot, time-lapse, and smartphone cameras has significantly increased the number of photos that are being captured on construction sites on a daily basis. Many photography documentation services have also emerged in recent years to deliver “visual” records of the as-built construction to project participants [21]. New aerial robotics companies are also emerging that can deliver large number of aerial as-built photos on a regular basis.

The research community has leveraged these still images together with multi-view geometry methods from the computer vision community for project-level monitoring purposes. Fig. 2 shows the overall concept of how still images can be used for progress and quality monitoring purposes. Time-lapse images are collected from fixed camera viewpoints to document the work-in-progress (WIP). These images are either compared with one another [6,22] or against a 4D BIM [23–28] which represents the expected state of construction progress. To highlight deviations in construction progress, several visualization methods are also proposed that color code construction elements based on the metaphor of traffic light colors [29]. Fig. 3 illustrate an example from [23,29] wherein a 4D BIM is superimposed on a time-lapse photo for progress monitoring purposes. Based on the metaphor of traffic light colors, the elements behind or ahead-of-schedule are color-coded with red and green colors respectively.

Another line of work automatically generates 3D as-built point cloud models of the ongoing construction using Structure from Motion (SfM) techniques and then compares the documented as-built model to the underlying 4D BIM. Golparvar-Fard et al. [23,31,32] and Brilakis et al. [33] conduct pioneer researches on SfM-based 3D as-built documentation. Golparvar-Fard et al. [2] improve the density of these 3D as-built point clouds by adopting a pipeline of Multi-View Stereo and voxel coloring algorithms, and present a method for superimposing point cloud models with BIM through a set of corresponding feature points between the as-built point cloud and BIM. In a more recent study, Karsch et al. [34] leverage BIM as a prior and presents a constrained-based procedure to improve the image-based 3D reconstruction. Their results show that the accuracy and density of image-based 3D reconstruction and backprojection of 3D BIM on unordered and un-calibrated site images can be improved compared to the state-of-the-art (see
Fig. 1. The workflow of recording, analyzing, devising, and implementing control decisions on many construction sites – data from BIMAnywhere and McCarthy construction based on a seminar presented at FIATECH2013.

Fig. 2. An overview of progress monitoring and quality assurance/quality control using still images. Image from [30].

Fig. 3. Superimposing 4D BIM on time-lapse images and color-coding progress deviations based on the metaphor of traffic light colors using the methods introduced in [23,29]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
example results in Fig. 4. More intelligent data collection methods such as [35,36] are also proposed that leverage aerial robots for collecting progress images. The underlying principles used in these methods for analyzing construction progress deviations roughly fall into two categories: (1) methods that reason about the physical occupancy in a scene from as-built models and (2) methods that infer changes in a scene mainly by observing appearance of the BIM elements in 2D images.

2.1. Occupancy-based assessment methods

Boscé et al. [37,38] introduce a method to use point clouds and BIM for monitoring construction of building elements on a project site. Building upon this method, Turkan et al. [1] use point clouds for tracking secondary and temporary concrete construction objects, for example by differentiating temporal objects such as formwork and rebars from a concrete column. More recently, Boscé et al. [5] introduce a method to use point clouds and BIM for monitoring Mechanical/Electrical/Plumbing (MEP) installations. A method introduced by Kim et al. [39] scans a building, matches the outcome with a BIM, and revises the as-built status. While these methods use laser scanning to generate point clouds, they are still applicable to point cloud models generated using image-based or video-based reconstruction procedures.

A method by Golparvar-Fard et al. [3] leverages integrated scenes of dense image-based 3D point clouds and BIM, and reasons about occupancy and visibility both from as-planned and as-built perspectives. With these assessments and a supervised machine learning method, progress is inferred for building elements in a binary fashion – i.e. physical progress is observed for an element or not. Fig. 5 illustrates the steps of creating 4D as-built point clouds, superimposing the point clouds with BIM, and automated progress monitoring from the generated DA4R – 4 dimensional augmented reality – models as introduced in [3,31].

Despite their importance, occupancy based methods are not capable of differentiating between different stages of operations involved in construction of an element. For example, with current accuracy for occupancy-based assessments, it is difficult to differentiate between forming, placing, and back-filling a concrete foundation wall and inferring the current state of progress. Han and Golparvar-Fard [40,41] attribute these challenges to (1) the lack of detail in BIM which is typically used as the basis for progress monitoring and (2) the lack of formalized construction sequences that allow operational details such as forming, reinforcing, placing, and stripping to be automatically inferred. Their work emphasizes the need for further research on appearance-based methods to differentiate and infer such details directly from the content of the 2D images. In the following, the state-of-the-art appearance-based methods are introduced and their applications in the context of automated progress monitoring from image-based point clouds and BIM are discussed.

2.2. Appearance-based methods

Over the past decade, several manual, semi-automated, and automated methods for progress monitoring using appearance information are proposed. A group of work focuses on the application of time-lapsed images and 3D models. Time-lapsed images are taken at a fixed location and are registered with BIM [23,25]. The registered images are analyzed and then compared with the as-planned model [23,25]. Similar approach is to observe significant visual changes among images taken at different points in time [24,26–28,42]. These changes are then compared to a 3D model for computing “plan percent complete”.

Another line of work focuses on detecting construction objects from site images. Zhu and Brilakis [43] propose using edge detection and Hough transform techniques for detecting concrete columns. Similarly, Wu et al. [44] propose an image segmentation method for detecting concrete columns. Kim et al. [45] use an installed pan–tilt–zoom camera and perform image processing procedures such as 3D CAD-based image mask filters and color-based noise removal, for updating a 4D CAD model. Kropp et al. [46] use distinct features found in different states of a drywall construction to monitor progress.

A main stream of research has focused on the broader problem of material classification from site images. Brilakis et al. [48,49], and Zhu and Brilakis [43] introduce image-based material classification techniques. Son et al. [50] compare the performance of different machine learning based algorithms that leverage color information for concrete detection. In Dimitrov and Golparvar-Fard [51], a Construction Material Library (CML) is introduced which contains 20 different material categories of 3000 image samples of materials, and a method for discriminative classification of construction materials from template images is presented. In a more recent line of work, Han and Golparvar-Fard [40,41,47] present several aspects of a new appearance-based material classification method for monitoring operation-level construction progress (see Fig. 6). Their method leverages 4D BIM and 3D point clouds generated from site images using Structure-from-Motion techniques. Through reasoning about occlusion, each BIM element is back-projected on all images which see that element. From these back-projections, several 2D image patches are sampled per element and are classified into different material types. By reasoning about the observed frequency of different material types, their method
is capable of tracking progress at the operational level beyond what is typically represented in a contractor’s schedule.

While significant progress in research is made on the fronts of occupancy and appearance-based methods, several key areas still remain open for future research. These open research challenges together are discussed in Section 4. In the following, the methods introduced for detecting, tracking, and recognizing activities of the construction equipment and workers and their applications are discussed in detail.

3. Visual monitoring of construction equipment and workers

Detection, tracking and analysis of construction entities to recognize their behavior follows the basic scheme of video analysis, which is also the pipeline of transforming visual information into applicable knowledge for construction monitoring at the operational level. Fig. 7 shows an overview of vision-based detecting, tracking, and analyzing activities of construction equipment and workers, as well as their applications in safety monitoring, occupational health assessments and carbon footprint benchmarking. In addition to regular detecting–tracking–analyzing pipeline, construction activities can also be analyzed by learning motion flow of feature points without explicitly tracking any entities [18].

3.1. Detection or recognition of construction entities

As mentioned above, detection or recognition of construction entities is the first step of vision-based monitoring at the operational level. It primarily involves detecting moving objects on the
construction sites or recognizing certain types of construction entities (e.g., workers or excavators).

Background subtraction is the most widely used method for moving object detection. It is assumed that a background image, which represents the static or relatively static scene, can be selected or generated from the video stream. Then given an input frame, moving objects can be detected by thresholding the difference between the background and the input images. Various methods have been presented and tested for modeling the background image imposing different statistical assumptions on the background pixels. Gong and Caldas [52] evaluate three methods (namely Mixtures of Gaussian, Codebook based method, Bayesian model based method) in construction scene and conclude that the Codebook based method and the Bayesian model based method outperform the Mixture of Gaussian. Chi and Caldas [53] adopt background subtraction to segment moving objects and classify workers, backhoes and loaders by several dimensional features and their average gray-scaled color. Notice that background subtraction is only applicable with a statically mounted camera and requires proper updating strategy to reflect the changes of the background scene. Since motion is the key factor, background subtraction is not applicable to detect still objects such as trucks or excavators in an idle state. Also, it tends to merge multiple objects into a single blob when they are adjacent to each other.

A general line of approach for construction entities detection/recognition is to exploit object shape features. This approach usually involves machine learning process. It generates a classifier through the training of the shape features. It starts with collecting image samples of various conditions in terms of viewpoint, illumination, and occlusion. Then feature descriptors are computed on the collected image samples. Two widely used feature descriptors are Haar-like features [54] and Histogram of Oriented Gradients (HOG) [55]. As for the machine learning process, Haar-like features are often combined with a cascade architecture – Adaboost (Adaptive Boosting) [54] which is also known as Haar Cascade, whereas HOG is commonly used with Support Vector Machine (SVM). Once classifiers are trained successfully, a multi-scale sliding window sweeps across the input image to locate the objects of interest.

In this pipeline, the feature descriptor is the key component affecting the detection performance. Researchers have adjusted cutting-edge visual descriptors for their specific purposes. Azar and McCabe [57] present two cascade approaches for dump truck recognition – Haar-HOG and Blob-HOG. Both approaches are comprised of two steps. In the first step, the former adopts Haar-like features with Adaboost while the latter uses background subtraction. In the second step, both approaches employ HOG based classifier. The first step plays a role of narrowing down the search region of HOG based classifier. To tackle with articulated objects, Azar and McCabe [58] adopt a part-based model using discriminatively trained HOG classifiers to detect excavators in different poses. Park and Brilakis [59] present a three-step procedure for construction equipment detection. In their method, images are processed with background subtraction, the Haar Cascade, and an eigen-image classifier which handle motion, shape and color information, respectively. Similarly, in [56], motion, shape and color features are used to detect construction workers wearing safety vests. The method first locates ordinary people using background subtraction and HOG based classifier, then distinguishes construction workers based on the color histograms trained with K-Nearest Neighbors. Accordingly, this method can be used to identify the safety vest wearing and facilitate onsite safety monitoring (Fig. 8). Instead of detecting construction entities separately with different features, a more recent line of research [60,61] presents a descriptor combining HOG and HOC (Histogram of Color) to detect excavators, trucks and workers simultaneously. HOG and HOC are formed, concatenated, and fed into multiple binary SVM classifiers to detect each type of resource separately. Their method leverages a non-maxima suppression algorithm on the sliding window candidates to accurately localize the detected resources based on the highest scores retrieved from the classifiers. This method is capable of detecting non-moving objects since it is not based on motion features. A comprehensive dataset of workers, excavators, and dump trucks for classifiers training is also proposed in their work. There are also several other attempts such as Weerasinghe and Ruwanpura [62] which leverage video and audio data for tracking workers and tools.

3.2. Tracking construction entities in 2D and 3D

Tracking in computer vision refers to locating moving objects in 2D or 3D across time. Compared to other tracking technologies such as GPS (Global Positioning System) [63], RFID (Radio Frequency Identification) [64] and UWB (Ultra-Wide Band) [65] (the latter two all require tagging every entity), vision based tracking has an advantage of its simple deployment. Different from detection, tracking finds the correspondence of entities across video frames.
and provides their trajectories. Visual tracking algorithms are often designed facing two issues: first, how an object is represented; second, how the correspondence of objects across video frames is established. Different solutions to these two issues, and various combinations between them result in tens of tracking algorithms. Generally, by the type of object representation, tracking algorithms can be divided into contour based, kernel based, points based and model based methods [66,67]. Researchers in construction focus less on developing new algorithms but on selecting a suitable one for their specific purposes. Hence many efforts have been dedicated to evaluate or adjust existing tracking algorithms in construction scenarios. Following is a brief overview of tracking algorithms alone. Applications of these algorithms in performance monitoring will be visited in the next subsection.

Teizer and Vela [20] compare four algorithms (namely mean-shift, Bayesian segmentation, active contours and graph cuts) on construction personnel tracking and conclude that Bayesian segmentation method performs the best. Similarly, Park et al. [68] compare kernel-based and point-based tracking algorithms on different types of construction entities (personnel, equipment and materials) and evaluate their performances on multiple metrics regarding illumination variation, occlusion and scale variation. Park and Brilakis [59] present a fusion of tracking and detection which enables automated initialization of tracking and enhances accuracy. It also exhibits the capability of handling occlusions as well as smooth tracking under viewpoint changes (Fig. 9). Brilakis et al. [69,70] propose to use multiple cameras to provide 3D trajectories of construction entities, which involves additional steps of camera calibration and triangulation. They compare the trajectory outputs to the total station data to evaluate the accuracy which is comparable to the GPS and their scheme is also experimented under harsh conditions to test its feasibility on real construction sites [71]. Yang et al. [72] develop a machine learning based scheme that can track multiple workers in a more efficient way. They compare the proposed system to the robotic total station (RTS) and conclude that the proposed visual tracking system is reliable and accurate enough for activity analysis purposes [73].

3.3. Activity recognition

With detection and tracking from previous stages, a higher level analysis is always expected to generate meaningful knowledge for various purposes such as operation analysis, proximity analysis for safety monitoring, occupational health assessments, and benchmarking and monitoring of carbon footprint for construction equipment. Fig. 10 shows simple scenarios of extracting useful information from tracking data. Tracking a concrete bucket and to check whether they are in safe locations all the time.

3.3.1. Operation analysis

Productivity measurement, cyclic operation analysis, and idle time detection are the main applications in operation level project monitoring. Various research works have been presented focusing on different types of entities and activities. Most heavily studied operation is in the earthworks of excavation, material hauling and dirt loading operated by hydraulic excavator and dump truck [60,74–76]. One of the earlier works is from Zou and Kim [74], which track the excavator by its distinguishable color from the surroundings in HSV color space and calculate its idle time by thresholding the centroid movement. Earthwork operations usually involve different types of equipment. To understand the complex activities, it is important to analyze the interaction among the entities. For example, dirt loading is usually operated by excavators and dump trucks. Therefore, both types should be investigated together to highlight their interactions. Azar et al. [76] detect and track an excavator and dump trucks simultaneously to estimate the dirt loading cycle.

The secondary is associated with the concrete placement activity which involves cyclic operations of a tower crane [77,78]. Gong and Caldas [77] detect a concrete bucket in video streams using the Haar Cascade and calculate its travel cycles based on the prior knowledge of construction site layout to extract productivity data. Yang et al. [78] perform similar work of monitoring concrete placement activity. They track the crane jib through 3D pose estimation. A probabilistic graph model is designed based on prior knowledge of the site layout to interpret the track signals and classify crane activities into two categories: concrete placement and non-concrete material movement (see Fig. 11).

Beside operation analysis of equipment, several studies dedicate to productivity evaluation or activity recognition of construction workers [52,79,80]. Peddi et al. [79] track workers tying rebar through blob matching, extract skeletons for pose estimation and classify their working status. Different from common video analysis pipeline, Gong et al. [18] adopt a bag-of-features strategy for worker activity recognition which directly extract spatial–temporal features from videos without explicitly tracking the workers. Golparvar-Fard et al. [17,81] also use spatio-temporal features together with multiple binary SVM classifiers to automatically label the activities of construction equipment in videos that capture a single piece of equipment (so called atomic activities). They also provide the research community with a large video database of atomic activities of excavators and dump trucks such as filling, hauling, dumping, swinging, moving, and idle. Their work in [82] also shows the application of atomic activity classifier on long sequences of videos where the equipment and activities are temporally localized (see Fig. 12).

Apart from processing videos captured by common cameras, adopting RGB-D cameras which bring in depth information has become a new trend. The main benefit here is that these methods

---

Fig. 8. Detection and classification of the construction workers: workers who are wearing safety vests (red) and not wearing safety vests (green) [56]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
bypass the more fundamental problem of detection and pose estimation which are necessary for activity recognition (see Fig. 13). These cameras provide a great opportunity for detecting and tracking workers in small scale indoor environments. Weerasinghe et al. [83] is one of the examples of these works for detecting and tracking the workers.

Escorcia et al. [84] present a method to recognize atomic activities of drywall construction. In their method, skeleton information is extracted from a sequence of RGB-D images (see Fig. 13) and multiple bag-of-poses are constructed and used to train discriminative activity recognition classifiers. Kim and Caldas [80] learn the workers action through Kinect supplied skeleton, train classifiers for their handheld tools (caulking gun, hammer, and electronic screw driver) and combine both the action and the tool type for activity recognition (caulking, hammering or screwing). Given a long sequence of RGB-D images, a new method developed by Khosrowpour et al. [19,85] automatically infers the time-series of activities conducted by workers involved in drywall operations (see Fig. 14). Their method divides the sequence into temporal segments and automatically classifies the observed actions. To do so, the algorithm first detects body postures in real time. Then a kernel density estimation (KDE) model is trained to model classification scores from discriminatingly trained bag-of-poses action classifiers from [84]. Further, a hidden Markov model (HMM) labels sequences of actions that are most discriminative. Their experimental results show the promise of the proposed method for deriving time-series of worker activities involved in drywall construction operations.

3.3.2. Proximity analysis and location tracking for safety monitoring

Safety is one of the critical factors to achieve success in construction projects. Common safety violations include speeding and proximity to predefined dangerous areas or heavy machinery. Real time 3D localization of construction entities is the core for safety assessment. Compared to RFID, radar, ultrasonic and GPS [86,87], video cameras are not dominant in this specific area but...
often used as assistant sensors for enhancement of environment perception. Chi and Caldas [88] use a stereo camera to track and recognize different types of equipment involved in earthmoving and report their speeding activities. Ishimoto et al. [89] mount a stereo rear monitor camera on an excavator to detect workers who appear within a certain range around the excavator.

Fig. 11. Analyzing cyclic operation of concrete placement by tracking tower crane through 3D pose estimation [78].

Fig. 12. Recognizing activities of construction equipment in long video sequences captured from earthmoving operations. Here, the equipment and atomic activities are temporally localized and the method introduced in [82] which primarily focuses on labeling these activities for carbon footprint monitoring purposes.

Fig. 13. Each column represents the depth image captured from a worker involved in drywall construction activities and the body posture that is automatically detected in real-time. This information is used in [84,85].
3.3.3. Occupational health assessments

Musculoskeletal disorders (MSDs) of workers are a serious issue in construction. Recent years, video cameras are utilized to acquire human movement footage and estimate joint loadings based on the whole body kinematics [90]. Unlike the wearable motion capture system (e.g. VICON, Xsense) or range cameras (e.g. Kinect [91]), it is difficult and computationally expensive to generate 3D motion skeletons from videos, which stays an open challenge in computer vision field. Limited attempts are either based on a calibrated monocular camera [92] or a stereo vision system [93–95]. Human intervention is required in all applications to manually identify the location of human joint centers for 2D skeleton extraction. Then 3D kinematic model is generated for further ergonomic analysis either by lumbar joint loading estimation [92] or classification into regular or unsafe motion based on machine learning methods [94,96]. For a comprehensive review on the application of computer vision methods for occupational health assessments, readers are encouraged to look into [90,96].

3.3.4. Benchmarking and monitoring of carbon footprint

Benchmarking and monitoring of carbon footprint are critical for minimizing environmental impacts in construction. Ideally sensors on board of the equipment would be used for measuring green house gas emissions. Nevertheless due to their high cost, research has focused on using computer vision techniques to understand the activities involved in an operation and then relating them into the expected and actual emission rates. Heydarian et al. [82] monitor carbon footprint of earthmoving operations using a vision based equipment activities recognition method along with pollutant emission inventories of construction activities. The monitored values are then compared to benchmark values to identify activities or a sequence of activities that could produce excessive pollutants. The introduced method can also be used to benchmark and monitor productivity and carbon footprint simultaneously and provide more insight as to how excessive emissions could be minimized.

4. Open research challenges and future directions

Although significant achievements are made in the past decade, yet a large amount of challenges still remain as open problems. Following the structure of previous literature review, several critical factors that currently hinder advancement for each component of the visual monitoring methods are discussed and possible immediate resolutions are introduced. Then a few long term challenges are presented for future consideration.

4.1. Roadmap for immediate research on visual monitoring of civil infrastructure or building elements

Research on leveraging still images for monitoring progress of civil infrastructure or building elements still has several open challenges. From a sensing perspective, the core of occupancy-based assessment methods is the 3D reconstruction of the built infrastructure. Most current efforts on generating point clouds still rely on laser scanners. Although image-based and video-based 3D reconstruction methods have shown very promising results, the accuracy and completeness of vision-based 3D reconstruction still need to be improved to minimize occlusions and provide reasonings under limited visibility. On the other hand, generating a binary decision of observed elements does not reveal the necessary information for control decision making. More details on the actual state of progress can be extracted from images. This requires the combination of occupancy-based and appearance-based methods. Recognizing objects in 2D images by color, texture, shape and other higher level information is quite a successful sub-field in computer vision. Future research should take advantage of object recognition in 2D and integrate with reconstructed 3D information for better observation of infrastructure or building elements. One area that needs further attention is the assembly of large datasets of construction material images that could be used for developing and benchmarking of future appearance-based recognition methods. Eventually, by generating semantically rich as-built BIM and comparing against as-planned BIM, project progress will be clearly and timely perceived. While the recent image-based dense reconstruction methods have shown very promising results, Ref. [97] suggests combining laser scanner and visual sensor may be a more realistic solution for reliable 3D reconstruction and detailed perception in short-term. In [98,99], a good example of combining laser scanner with video cameras has been given. The challenges of improving as-planned BIM and the underlying Work Breakdown Structure as well as formalizing knowledge of construction sequencing are other open areas that still require further research.
4.2. Roadmap for immediate research on vision-based detection, tracking, and activity recognition of construction resources

The methods developed for progress and quality monitoring often leverage strong priors such as a 3D Building Information Model (BIM) and schedule which provide detailed information about the geometry and appearance of the elements in any point in time. In contrary, the methods used for tracking and analyzing activities of the equipment and labor may not have any strong priors. The large inter and intra-class variability among these resources from one side, and the lack of publicly available datasets for creating new algorithms and benchmarking their performance are two major barriers that have slowed the development of these methods.

For construction entity detection, it can be concluded from the literature review that only limited types of entities have been studied, including workers, excavators, dump trucks, loaders and tower cranes. Apparently there are still many other types of construction equipment to be explored, and multiclass detection is not soundly established. Furthermore, researchers tend to train and test classifiers on their own dataset. Sharing the dataset will benefit the community by facilitating direct comparison and evaluation of the varied methods. Detection performance across different scenarios is also unknown yet. Though construction sites are highly dynamic, cluttered and project oriented, the entities across different sites and projects share many common features due to the standardization of construction procedures. Similar to the built-in face detection function for digital cameras or automatic pedestrian detection for smart cars, it is helpful to build up standard image library covering key types of construction entities for training more generic resource classifiers. Moreover, different types of equipment are usually for specific tasks. For example, excavators and dump trucks are usually related to earth moving operations. Multiclass entity detection can be simplified to some extent if the construction task being executed is given as a prior knowledge.

Tracking itself is still an open issue in computer vision field. Occlusion (static or dynamic), articulated or non-rigid object and view/scale/illumination changes all pose big challenges to tracking. These situations can only be worse in cluttered construction sites. Besides these common issues, some unique characteristics of construction projects prevents the direct use of existing tracking methods. First, scenarios are highly dynamic in construction sites. Not to mention that the project progress is making changes to the site every day, in a shorter duration, the moving equipment and leaving and re-entering entities will not only challenge background modeling based tracking, but cause troubles for identity preserving of long-term tracking, eventually cause troubles in the subsequent activity analysis. Secondly, the appearance similarity of construction equipment and workers leads to difficulties for appearance based tracking, especially for multiple interacting workers tracking. Thirdly, the complex motion of construction entities often combines the modes of locomotion and joint motion. For example, an excavator drives to a spot and starts excavating. Or a worker lifts rebars to somewhere and begins tying them. This will definitely challenge existing tracking methods which are designed for a single mode of motion. Fourthly, interaction, either between workers or workers and equipment, happens frequently in construction project, which is another big issue for multiple objects tracking. Hence, long-term stable tracking of construction entities in real time is still far from being realized. The duration of a stable tracking at construction sites is at most a few minutes long, which is totally incomparable to the natural length of hours of videos. This limits the practical use of the visual tracking in construction sites. Though previous studies contribute to take a significant step forward in introducing computer vision technology, most of them adopt relatively old algorithms. There are many other good alternatives. For example, TLD (Tracking–Learning–Detection) [100] is a very stable long term tracker. Combining with online learning and detection is an inexorable trend for tracking. From a long term perspective, it is reasonable to use visual tracking parallel with other types of sensing technologies, including GPS [63], UWB [65], and RFID [64]. The various strengths and drawbacks of different systems will compensate each other. Non-visual sensor based tracking is more robust to visual occlusion, multi-objects interaction and environmental changes while visual tracking can capture more details about the actions (e.g. body gestures), as well as semantic information from the surroundings for better activity understanding.

Construction is a well-planned, complicated process containing various activities. Existing studies explore a very small sets of construction activities, mainly on earthmoving, concrete placements, drywall construction, and several simple worker activities. Researchers usually take the type of the activity as a prior knowledge and analyze its cyclic operation flow. Each type of activities is treated as an independent procedure and isolated from the entire construction process. Holistic understanding of the scene can be used to improve the situation. Lots of prior known information is available from the construction planning stage, such as BIM, site layout, construction schedule, etc. By combining these prior knowledge, not only activity understanding is enhanced, but also systematic project controls can be realized. Furthermore, it is known to all that most computer vision algorithms are computationally expensive. Thus it is still difficult to analyze all the recorded construction videos. A sampling strategy can be introduced to extract short clips from long videos. Samples may be taken according to the importance of various tasks and their time dependency as to the project schedule. Then the productivity information or progress deviations can be inferred based on a smaller video set. One possible solution for developing more enhanced methods for activity recognition of construction workers could be devising strategies that can involve construction companies together with non-expert annotators to collect very large datasets in form of compositional structures of activities. For example, creating datasets that encapsulate the posture and tools used in any given atomic activity conducted by the workers involved in specific construction roles (vibrator, finisher, pump operator, helper, etc. in concrete placement operations).

4.3. Roadmap for future research

Apart from the aforementioned challenges, some future possible directions may include the following scenarios:

(1) Construction is the transition process from input (labor, equipment, materials) to output (built infrastructures). Monitoring infrastructure elements is an assessment from the output side, while monitoring operations of equipment and workers is an observation from the input. Fusing information from both sides may give comprehensive knowledge on ‘how a task is operated’ (workflow, productivity), ‘how much work is done’ (progress) and ‘how well it is finished’ (quality). This will definitely lead to better understanding/evaluation of the construction process. The recent work in [101] presents a good working scenario, wherein an earth-moving operation is monitored by both measuring the volume of excavated soil and analyzing the activities conducted by the excavators.

(2) The future of intelligent construction monitoring can rely on multiple types of sensors or sensor networks. Combining with other types of sensors (such as laser scanner, GPS, RFID, etc.) will compensate the disadvantages of visual sensors and provide more types of information. Data fusion from
multiple resources should be studied. Integrated comparison with other types of sensors is also required to clearly investigate the pros and cons of the technologies in terms of entity types, site conditions, construction activities, and coverage areas.

(3) Gap exists between research and industrial application. To promote widespread adoption, research should be more generalized and user interfaces should be built up for configurations towards different situations. The problem of artificial intelligence for construction often requires adaptations and specialized techniques. To embed intelligence, codifying much of what goes on in the practice is needed at a larger scale.

Ultimately the purpose of construction performance monitoring is to provide information on the most updated state of ongoing operations so that the practitioners can “control” potential or actual poor performances or deviations. Nevertheless, research in monitoring performance is not mature enough to enable control related studies. A foreseeable future is that construction progress can be analyzed and project schedule can be revised automatically through observation from the visual sensors. Or for example, productivity can be estimated and operation procedures can be optimized based on the analysis of the current activities. All these require the inclusion of artificial intelligence or machine learning into the solution framework. Adopting advanced, cutting-edge computer vision algorithms is always the first consideration. However, the experience of construction researchers must be involved for meaningful application. Prospering future of this very exciting field with great practical significance requires intimate collaborations between computer vision and construction researchers.

5. Conclusion

This paper provides a detailed review of the literature on computer vision methods for monitoring construction performance both at project level and operation level. A detailed assessment of these methods and their assumptions is provided. Most of these methods are still at their early stages of the development, thus the focus was primarily placed on comparing their accuracy and applicability to specific problems in performance monitoring. A roadmap for immediate and future research on visual monitoring of civil infrastructure elements and also construction operations is also proposed. No specific discussion on computational time was presented because the presented methods are primarily developed to validate early concepts and software optimization for minimizing computation time in most parts is still under development. Since the codes for implementing these methods are also not provided and there is a lack of comprehensive datasets for validations, the comparison of these methods has remain as part of the future work.

Acknowledgements

Jun Yang is supported by National Natural Science Foundation of China No. 51208425 and Research Foundation of Northwestern Polytechnical University No. JCY20130127. Man-Woo Park is supported by the Grant No. 14CTAP-C078828-01 from Infrastructure and Transportation Technology Promotion Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government. Mani Golparvar-Fard is supported by a faculty fellowship from the United State National Center for Supercomputing Applications and also by a grant from the Institute of Applied Science and Technology (ICTAS). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of organizations mentioned above.

References


