

Deep learning-based Context Documentation for Earthmoving Productivity Simulation

Hongjo Kim[a,b], Ph.D., Youngjib Ham[a], Ph.D., and Hyoungkwan Kim[b], Ph.D.
[a]Department of Construction Science, Texas A&M University, Texas, United States
[b]School of Civil and Environmental Engineering, Yonsei University, Seoul, Republic of Korea

Although construction process simulation has shown its potential for analyzing earthwork productivity, the time-consuming and laborious data collection process remains as a major obstacle to implement it in practice. To address this issue, this paper proposes a deep-learning based context documentation method for earthmoving productivity simulation. The proposed method automatically produces the site access log of dump trucks using a visual sensing process consisting of video deinterlacing, convolutional networks, and context reasoning. The site access log is then used to construct simulation inputs—the durations of each earthmoving task—in the form of probabilistic distributions. Using the obtained simulation input data, the sensitivity analysis of construction process simulation provides an optimal resource allocation plan in a daily basis. A case study was conducted to validate the proposed method and demonstrated its applicability in real construction sites. It is expected that the proposed method will facilitate accurate decision-makings in earthmoving resource allocation based on real construction operational data.

Key Words: Earthmoving; Productivity Analysis; Resource Allocation; Nonintrusive Monitoring; Computer Vision; Simulation

Introduction

In construction projects, various earthworks are generally involved with the cutting, filling and moving soils and rocks using heavy vehicles such as dump trucks, excavators, bulldozers and compactors. Since earthworks comprise a large portion of construction projects in terms of costs and schedules, researchers have long been investigated to improve the earthwork productivity by optimizing resource allocation (Parente et al. 2015), schedules (Hwang et al. 2014), or work sequences (Gwak et al. 2018). Although such previous studies have proposed valuable earthmoving analysis methods including construction process simulation that can yield the promising productivity improvements, the construction industry still depends on empirical decision making to arrange earthwork operations (Gao et al. 2014), especially in outdoor construction sites.

Regarding this situation, two main reasons were discerned through literature review and interviews with experts in the construction industry: First, construction practitioners prefer to plan earthwork operations based on their experience due to hectic schedules and the lack of available budgets to analyze the earthwork productivity. Second, an earthwork plan established by a laborious analytic process is likely to become obsolete in a short time due to the changing geologic conditions and the operational situation of construction projects. Owing to these issues, the implementation of existing earthwork analysis methods has been hindered in the construction industry.

Automating the data collection process for earthwork analysis methods has a potential for solving these issues, since it has been known as one of the most time-consuming and laborious tasks for earthwork productivity analysis. As for construction process simulation, such as CYCLONE (Halpin 1977), SIMPHONY (Hajjar and AbouRizk 1999), and STROBOSCOPE (Martinez and Ioannou 1994), the context of earthworks—the durations of earthwork tasks—should be frequently updated to reflect the actual earthwork performance, thereby ensuring the reliability of a simulation model and its results (Akhavian and Behzadan 2015; Vahdatikhaki and Hammad 2014). In this respect, vision-based monitoring methods have the benefits of automating data collection for earthwork operations, as they demonstrated their applicability in various jobsite monitoring applications (Bang et al. 2017; Hamledari et al. 2017; Kim et al. 2016; Kim et al. 2017).

However, there are underlying challenges in realizing the benefits of the vision-based monitoring methods in outdoor construction sites for earthwork productivity analysis. First, a complex multi-camera monitoring system is necessary to monitor large earthwork areas. Accordingly, proper image analysis methods need to be developed to deal with a large amount of image data, multi-camera calibration, and the topology of camera views (Wang

2013). Second, project-related confidentiality or worker's privacy issues may preclude vision-based monitoring for the earthwork productivity analysis.

To address the challenges, this study presents a novel way of automating the earthwork data collection for construction process simulation, which employs indirect monitoring of earthwork operations using a computational framework consisting of video pre-processing, computer vision, and a context reasoning module for earthwork operations, as shown in Figure 1. Through the proposed method as shown in Figure 1, the site access log of dump trucks is obtained from jobsite entrance videos by visual analytics which include video recording, video deinterlacing, and LP detection and LP recognition; then, the site access log is utilized in the earthmoving event reasoning to infer the context of an earthmoving process in which multiple excavators and dump trucks interact with each other. Here, the context is represented as the time durations of earthmoving tasks. As a result, the durations of each earthmoving task are automatically obtained, and they are used to construct a simulation input for simulation. To produce an earthmoving productivity report, a sensitivity analysis regarding different numbers of earthmoving resources—excavators and dump trucks—is performed. The underlying research question in this paper is as follow: "Can the vision-based context documentation at a site entrance and exit provide reliable contextual information regarding earthmoving processes without monitoring real operations?". To answer this research question, this study conducts a case study at a real construction site. The main contributions of this study include (1) introducing a novel earthmoving productivity analysis method which integrates nonintrusive monitoring and simulation methods, (2) presenting how to achieve the high performance of LP detection and recognition in an uncontrolled environment, and (3) demonstrating the reliability of the resulting simulation outputs, which helps data-driven decision-making in earthmoving resource allocation. The following sections elucidates relevant previous studies, the methodological details of the proposed method, the case study results, and the discussion.

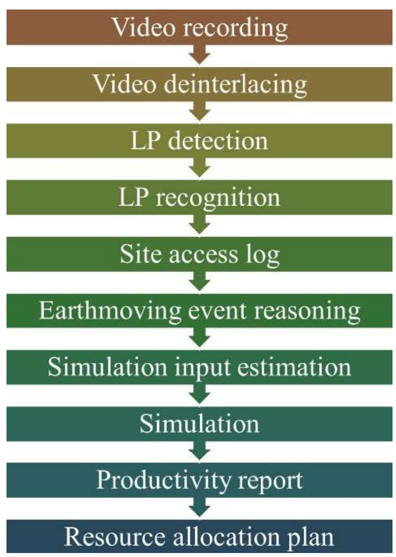

Figure 1. Overview of the proposed earthmoving productivity analysis

Related Works

Discrete event simulation (DES), such as CYCLONE (Halpin 1977), SIMPHONY (Hajjar and AbouRizk 1999), and STROBOSCOPE (Martinez and Ioannou 1994), has been used in previous studies for analyzing earthwork productivity because of its advantage in handling uncertainties. It considers uncertainties in the form of probabilistic distributions representing the durations of construction work tasks (AbouRizk 2010). However, in practice, DES has not been widely used for optimizing earthwork productivity due to the cumbersome data collection process for constructing simulation inputs (AbouRizk et al. 2011). To overcome this issue, various methodologies have been investigated to automate jobsite data collection processes for enabling the generation of contextual information.

A few studies proposed analytic methods to produce jobsite contextual information regarding earthmoving processes from images (Bügler et al. 2017; Kim et al. 2018a; Rezazadeh Azar et al. 2013). The principle of their methods is first to detect construction objects such as dump trucks and excavators in images, and then measure the durations of each work task through the rule-based context reasoning to record a jobsite activity log. Based on the activity log, the parameters of probabilistic density functions that are used in construction process simulation can be formulated using density estimation methods such as maximum likelihood estimation (Kim et al. 2018a). However, as discussed in the introduction, the necessity of a complex video surveillance system or worker's privacy issues may hinder their methods to be implemented in practice, especially for outdoor construction sites.

As an alternative, indirect monitoring of construction operations can be used to infer jobsite contextual information; such monitoring methods do not directly monitor jobsite activities by recording construction operations, but, for example, record the access of construction vehicles such as dump trucks when they pass through a site entrance. Therefore, the challenges of vision-based monitoring (e.g., the necessity of a complex multi-camera monitoring system, project-related confidentiality, or worker's privacy issues) are no longer involved. If the site access log contains each vehicle identity with its access time, the durations of earthwork tasks can be estimated based on a proper context reasoning module.

To automate the documentation of the site access log, video data containing the site access information of construction vehicles is a requisite; such videos can be obtained by a closed-circuit television (CCTV) or a camcorder. To the authors' best knowledge, the video surveillance system recording a jobsite entrance is generally used in many construction projects for security reasons. This study aims at utilizing such available resources to analyze earthwork productivity, specifically for an earthmoving process. In this respect, an automatic license plate recognition (ALPR) is appropriate to document the site access of construction vehicles. However, jobsite monitoring characteristics, such as multi-directional appearances, illumination variations, and low spatial resolutions, should be considered for reliable ALPR since small errors in recognition may lead to a large deviation in productivity analysis results (Kim et al. 2018a). To secure high accuracy in ALPR, various studies utilized deep learning-based object detection models. Deep learning denotes a learning process of specific models which have hierarchical deep architectures (Kim et al. 2018b). Regarding visual tasks such as object classification and detection, convolutional neural networks have been widely used to employ their capability of extracting hierarchical visual features of target objects based on training image data. Nevertheless, the deep learning-based ALPR models are likely to yield inaccurate performance due to the viewpoint variations (Xie et al. 2018), low spatial resolutions (Chen et al. 2016), and motion blur in uncontrolled jobsite environments.

Methodology

Among various earthwork processes, an earthmoving process is selected to demonstrate the proposed method. To produce simulation input data, the site access log is generated by detecting license plates and their numbers from videos collected from a site entrance. Since this study uses a video dedicated to monitoring a jobsite entrance for security reasons, undesirable issues would be involved in image data contents, such as low spatial resolutions and blurriness, multi-directional appearances, motion blur, irregular shapes of target objects, and illumination variations. To obtain high performance in license plate detection and recognition, videos are preprocessed to reduce motion blur, and then deep convolutional networks detects the license plates and their numbers in dump trucks to deal with the remaining issues. Based on the detection results, the durations of each earthmoving task are estimated, and they are used as an input for construction process simulation—WebCYCLONE. The following subsections describe the details of the proposed method from video deinterlacing to simulation, as shown in Figure 1.

Video deinterlacing for improving the visibility of license plates in jobsites

At the jobsite entrance, a camcorder records the site access of dump trucks passing through without stopping. The movement of dump trucks leaves motion blur in video frames along with the flow of their motion. Motion blur in video frames significantly degrades the detection performance for license plates and numbers, since their appearance become vague in such situations as shown in Figure 2. To reduce motion blur in video frames, a video deinterlacing technique by linear interpolation is used, which adjust each pixel value based on its adjacent pixel values. For example, when applying video deinterlacing by linear interpolation with respect to a video frame, a pixel in odd pixel lines has its pixel value by averaging two adjacent pixel values at even pixel lines. For the next video frame, the same operation is applied to the even pixel lines. The right part of Figure 2 shows a result of video deinterlacing.

Figure 2. Effects of video deinterlacing (Left: before deinterlacing, Right: after deinterlacing).

License plate detection for dump trucks

Although motion blur can be removed by video deinterlacing, the remaining challenges still need to be addressed to obtain accurate detection results regarding license plates and their numbers. This study employs a convolutional network and customize it to detect the target objects, building on a region-based fully convolutional networks (R-FCN) (Dai et al. 2016). In general, convolutional networks require a large amount of training data for each class to be trained, otherwise the overfitting problem impairs the detection accuracy. To alleviate the overfitting problem, transfer learning is applied to train the R-FCN model. In transfer learning, a pre-trained model using a large amount of training data in different domains is re-trained with the relatively small amount of training data in this study. By doing so, the robust detection capability of convolutional networks is maintained, thus, the reliability of the earthmoving contextual information can be improved. To deal with varying sizes of target objects, the range of anchor boxes, which are candidate regions likely to have the target objects—is changed from 128^2, 256^2, 512^2 pixels to 8^2, 16^2, 32^2, 64^2, 128^2, and 256^2 pixels. Moreover, to further reduce the overfitting issue, training image data is augmented by changing pixel's intensity value in each color channel, or adding frequency noises, contrast, Gaussian blur, average blur, and median blur, as shown in Figure 3.

Figure 3. Examples of image augmentation

License plate recognition and simulation input estimation

The detection results include the presence of license plates and plate numbers in each image. With the video frame rate of 30, hundreds or thousands of video frames are recorded within a short time, which have the same dump truck. In some cases, the detection results may vary from other video frames, and this causes an error in measuring the durations of earthmoving tasks. To secure the high accuracy, three rules are implemented in the license plate recognition algorithm, as shown in Figure 4: (1) performing the detection of license numbers only when the width of a detected license plate exceeds 150 pixels; (2) documenting license plate numbers when the number of detected license plate numbers is exactly seven; and (3) filtering out the license plate numbers when these are not matched with registered license plate numbers of dump trucks. Based on these processes, the reliable site access log can be produced. The site access log is then used to measure the time durations of earthmoving tasks for constructing an input—parameters of probabilistic distributions—for construction process simulation.

```
INPUT: Nth video frame
OUTPUT: license plate numbers
FOR N = 1:end of video frames
    Perform object detection with R-FCN
    IF width of a license plate > 150 pixels & number of detected objects == 7
        Order the detected plate numbers from left to right
        IF Any of registered plate numbers is same as the detected plate numbers
            Record the plate numbers with the video frame name
        END
    END
END
```

Figure 4. Algorithm for license plate recognition

Site access log-based productivity simulation analysis

The site access log includes the arrival and departure time of each dump truck at the site entrance, and this information is used to estimate the durations of a loading, hauling, and retuning tasks. An interval between the arrival and departure time of the same dump truck is used to measure a single loading duration. Likewise, an interval between the departure and arrival time of the same dump truck is used to measure a hauling duration and a returning duration, by dividing the time of the round trip by two to record each duration. WebCYCLONE (Web-Cyclic Operation Network, Halpin et al. (2003)) is selected as a simulation tool for the earthmoving process, since it has been validated by many studies regarding productivity analysis in the construction domain (Halpin et al. 2003). An earthmoving simulation model is designed to include two COMBI elements for loading and dumping tasks, and two NORMAL elements for hauling and returning tasks. The durations for these elements are constructed as beta distributions by the method-of-moments (Bain and Engelhardt 2000), which estimates the shape parameters of a beta distribution based on the given data—the site access log, as follows:

$$\alpha = \bar{x}\left(\frac{\bar{x}(1-\bar{x})}{\bar{v}} - 1\right) \tag{1}$$

$$\beta = (1-\bar{x})\left(\frac{\bar{x}(1-\bar{x})}{\bar{v}} - 1\right) \tag{2}$$

where $\bar{x}$ is the sample mean and $\bar{v}$ is the sample variance.

Experimental results and discussion

The experiments were performed using a normal desktop with Intel i7-6700 CPU and a GeForce GTX 1080 in the Ubuntu 16.04 operating system. As a case study, an actual earthmoving process at a construction site in South Korea was investigated, where four excavators and eighteen dump trucks were interacting with. A SONY HDR-CX450 camcorder was used to record the site entrance video with the resolution of 1920×1080 for an entire day. The proposed context estimation method was tested on this video, and the precision of 97.06% and the recall of 92.94% were recorded in license plate recognition. As shown in Figure 5, the proposed method successfully identified license plates and numbers in the uncontrolled environment. Without video deinterlacing, the performance was significantly degraded to the precision of 90.48% and the recall of 40.71%. These results imply the importance of video deinterlacing in monitoring at a construction site entrance, and the applicability of deep convolutional networks in detecting varying sizes of objects in an uncontrolled environment. Based on the license plate recognition results, the daily site access log was produced, and it was used to estimate the shape parameters of beta distributions, resulting in α of 0.750 and β of 2.345 for loading tasks, α of 1.328 and β of 2.490 for hauling and returning tasks.

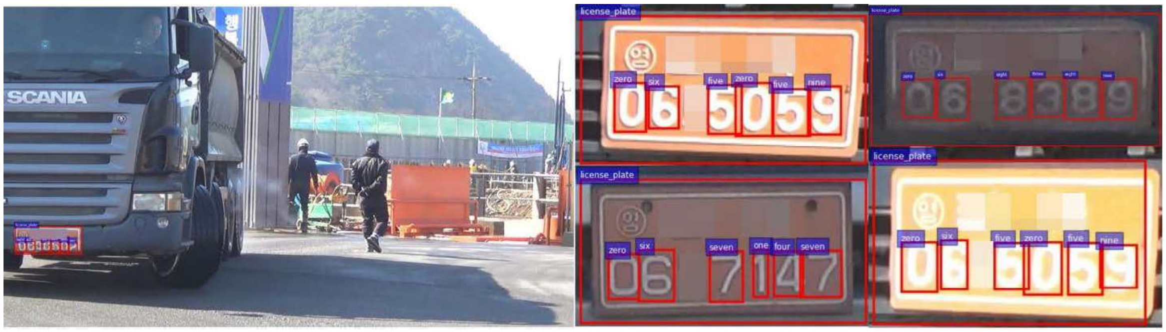

Figure 5. Examples of the license plate detection results (Left: a detection result in a video frame, Right: detection results enlarged at license plates).

At the case study site, the total planned earthmoving volume was 258,047m^3 with four excavators and eighteen dump trucks. The validity of a simulation model was tested by comparing the simulation results with the actual earthmoving performance. The difference was only 3.24% with respect to the time spent for the same number of earthmoving cycles, which proved that the simulation model was valid. After performing simulation, the current resource allocation was identified to be suboptimal, recording the idle time of 68.75% for excavators and 13.34% for dump trucks, as shown in Table 1. Based on the comprehensive list of productivity indices as shown in Table 1, site managers can compare the expected results with different numbers of dump trucks. Since the required information to optimize the earthmoving process varies depending on situations, site managers may set certain criteria to allocate earthmoving resources using one or more of these productivity indices. In this study, the Unit cost ($/m^3) is set as the main criterion for the resource allocation. By reducing the idle time of excavators while using 24 dump trucks, the earthwork productivity is expected to be improved such that both the original schedule and cost are reduced by 23.94% and 3.47%, respectively. Although an amount of reduced costs is not that significant when 24 dump trucks are used, it is expected that the reduced schedule may also reduce associated indirect costs in earthmoving operations.

Table 1. Simulation results with a different number of dump trucks

Number of dump trucks	18	24
Process cycle per hour (cycle/h)	12.60	16.56
Productivity per hour (m^3/hour)	214.2	281.5
Rental cost per day ($/day)	12,442	15,785
Total production of a day (m^3/day)	1820.7	2392.9
Unit cost ($/m^3)	6.83	6.60
Idle time of excavators (%)	68.75	34.86
Idle time of dump trucks (%)	13.34	41.46
Estimated earthmoving period (day)	142	108
Estimated total earthmoving cost ($)	1,763,395	1,702,174

Conclusion

The construction industry has been suffered from the inefficiency of data collection and communication due to the unique project-based nature, thus, it is difficult to make data-driven decisions in various aspects. This study dedicated to address this issue by automating the data collection process in earthmoving processes based on vision-based monitoring techniques. Moreover, construction process simulation was integrated with the vision-based monitoring system, to further automate the process of productivity analysis in earthmoving operations. The proposed method demonstrated its capability in this respect, building on the deep-learning based earthmoving context reasoning and the WebCYCLONE simulation. Despite the applicability of this method in an earthmoving process, further study is required to increase its generalization capability for the dissemination of vision-based productivity analysis methods in various stages during construction projects. One area of further studies is to improve the visual recognition performance for target objects having insufficient geometric features. For example, it was relatively hard to detect the object class “one” due to its monotonous shape. In jobsites, there are many objects having less geometric features such as steel beams, therefore, it would be helpful to investigate a method to robustly detect such target objects. In addition, the current study only analyzed a single earthmoving construction process, but further study needs to be conducted to increase its capability in analyzing multiple construction operations at the same time. Meanwhile, various external factors affecting the productivity of construction operations should be also considered to realize a robust data-driven decision support system.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science, ICT and Future Planning, No. 2018R1A2B2008600; Ministry of Education, No. NRF-2018R1A6A1A08025348) and the Yonsei University Research Fund (Yonsei Frontier Lab. Young Researcher Supporting Program) of 2018.

References

AbouRizk, S. (2010). "Role of simulation in construction engineering and management." *Journal of construction engineering and management*, 136(10), 1140-1153.

AbouRizk, S., Halpin, D., Mohamed, Y., and Hermann, U. (2011). "Research in modeling and simulation for improving construction engineering operations." *Journal of Construction Engineering and Management*, 137(10), 843-852.

Akhavian, R., and Behzadan, A. H. (2015). "Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers." *Advanced Engineering Informatics*, 29(4), 867-877.

Bain, L. J., and Engelhardt, M. (2000). *Introduction to Probability and Mathematical Statistics*, Duxbury Press.

Bang, S., Kim, H., and Kim, H. (2017). "UAV-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching." *Automation in Construction*, 84, 70-80.

Bügler, M., Borrmann, A., Ogunmakin, G., Vela, P., and Teizer, J. (2017). "Fusion of Photogrammetry and Video Analysis for Productivity Assessment of Earthwork Processes." *Computer‐Aided Civil and Infrastructure Engineering*, 32, 107-123.

Chen, C., Liu, M.-Y., Tuzel, O., and Xiao, J. "R-CNN for small object detection." *Proc., Asian Conference on Computer Vision*, Springer, 214-230.

Dai, J., Li, Y., He, K., and Sun, J. (2016). "R-FCN: Object detection via region-based fully convolutional networks." *Proc., Advances in Neural Information Processing Systems*, Neural Information Processing Systems Foundation, Barcelona, Spain, 379-387.

Gao, T., Ergan, S., Akinci, B., and Garrett, J. H. (2014). "Proactive Productivity Management at Job Sites: Understanding Characteristics of Assumptions Made for Construction Processes during Planning Based on Case Studies and Interviews." *Journal of Construction Engineering and Management*, 140(3), 04013054.

Gwak, H.-S., Seo, J., and Lee, D.-E. (2018). "Optimal cut-fill pairing and sequencing method in earthwork operation." *Automation in Construction*, 87, 60-73.

Hajjar, D., and AbouRizk, S. "Simphony: an environment for building special purpose construction simulation tools." *Proc., the 31st conference on Winter simulation: Simulation---a bridge to the future-Volume 2*, ACM, 998-1006.

Halpin, D. W. (1977). "CYCLONE-method for modeling job site processes." *Journal of the Construction Division*, 103(3), 489-499.

Halpin, D. W., Jen, H., and Kim, J. "A construction process simulation web service." *Proc., the 35th conference on Winter simulation: driving innovation*, Winter Simulation Conference, 1503-1509.

Hamledari, H., McCabe, B., and Davari, S. (2017). "Automated computer vision-based detection of components of under-construction indoor partitions." *Automation in Construction*, 74, 78-94.

Hwang, S.-I., Son, J.-H., and Lee, S.-H. (2014). "Development of scheduling model for earth work using genetic algorithm." *KSCE J. Civ. Eng.*, 18(6), 1618-1624.

Kim, H., Bang, S., Jeong, H., Ham, Y., and Kim, H. (2018a). "Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation." *Automation in Construction*, 92, 188-198.

Kim, H., Kim, H., Hong, Y. W., and Byun, H. (2018b). "Detecting Construction Equipment Using a Region-based Fully Convolutional Network and Transfer Learning." *Journal of Computing in Civil Engineering*, 32(2), 04017082.

Kim, H., Kim, K., and Kim, H. (2016). "Data-driven scene parsing method for recognizing construction site objects in the whole image." *Automation in Construction*, 71, 271-282.

Kim, K., Kim, H., and Kim, H. (2017). "Image-based construction hazard avoidance system using augmented reality in wearable device." *Automation in Construction*, 83, 390-403.

Martinez, J. C., and Ioannou, P. G. "General purpose simulation with stroboscope." *Proc., Proceedings of the 26th conference on Winter simulation*, Society for Computer Simulation International, 1159-1166.

Parente, M., Cortez, P., and Gomes Correia, A. (2015). "An evolutionary multi-objective optimization system for earthworks." *Expert Syst. Appl.*, 42(19), 6674-6685.

Rezazadeh Azar, E., Dickinson, S., and McCabe, B. (2013). "Server-Customer Interaction Tracker: Computer Vision–Based System to Estimate Dirt-Loading Cycles." *Journal of Construction Engineering and Management*, 139(7), 785-794.

Vahdatikhaki, F., and Hammad, A. (2014). "Framework for near real-time simulation of earthmoving projects using location tracking technologies." *Automation in Construction*, 42, 50-67.

Wang, X. (2013). "Intelligent multi-camera video surveillance: A review." *Pattern Recognition Letters*, 34(1), 3-19.

Xie, L., Ahmad, T., Jin, L., Liu, Y., and Zhang, S. (2018). "A New CNN-Based Method for Multi-Directional Car License Plate Detection." *IEEE Transactions on Intelligent Transportation Systems*.