

VIDEO ANALYSIS FOR TOWER CRANE PRODUCTION RATE ESTIMATION

SUBMITTED: August 2022
REVISED: November 2022
PUBLISHED: February 2023
EDITOR: Bimal Kumar
DOI: [10.36680/j.itcon.2023.007](https://doi.org/10.36680/j.itcon.2023.007)

El-Badr O. Elgendi

*Construction and Construction Department,
Arab Academy for Science, Technology and Maritime Transport Alexandria, Egypt
Email: elbadrosman@aast.edu*

Khaled M. Shawki

*Construction and Construction Department,
Arab Academy for Science, Technology and Maritime Transport Alexandria, Egypt
Email: kshawki.eng@gmail.com*

Amr Ashraf Mohy

*Construction and Construction Department,
Arab Academy for Science, Technology and Maritime Transport Alexandria, Egypt
Email: a.el-deen5146@student.aast.edu*

SUMMARY: Construction equipment production rates are considered an influential factor in construction projects' success as 30–40% of construction cost overruns are attributed to insufficient equipment production rates. The tower crane is one of the heavy construction equipment, and its production rate has a huge impact on the construction project's performance. Current estimations of the tower crane production rate on construction sites are not backed by a universally accepted and applicable methodology. However, vision-based technologies have been utilized recently for determining the production rates of construction equipment. Therefore, the purpose of this research is to develop a vision-based research framework (VRF) with a user-friendly interface for a practical and speedy measure of tower crane actual cycle time on construction sites. The software will be developed based on eight single object tracking detection free algorithms. Moreover, the VFR will be evaluated by measuring tower cranes' cycle time in two case studies in Egypt. For each case study, five videos were recorded for the tower cranes. The cycle time in each video was measured by using manual inspection and by the VRF. The VRF achieved high accuracy in tracking the cycle time for the tower crane in the two case studies.

KEYWORDS: Tower crane, video analysis, production rate analysis, Construction equipment, programming language, Vision-based technologies

REFERENCE: El-Badr O. Elgendi, Khaled M. Shawki, Amr Ashraf Mohy (2023). Video analysis for tower crane production rate estimation. *Journal of Information Technology in Construction (ITcon)*, Vol. 28, pg. 138-150, DOI: [10.36680/j.itcon.2023.007](https://doi.org/10.36680/j.itcon.2023.007)

COPYRIGHT: © 2023 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



1 INTRODUCTION

The cost of construction equipment might range from 10 to 40 percent of the entire project cost; therefore, equipment production rates are considered an influential factor in projects' success, especially with the huge reliance on heavy equipment in construction nowadays (Alaghbari et al., 2019). However, it is generally acknowledged that work in the construction industry is performed under exceedingly challenging site conditions due to the incorporation of heavy equipment into the processes of construction, such as tower cranes, with the continuous movement of workers, materials, and changing status (Kim and Caldas, 2013). Tracking construction projects and giving an accurate estimate of the construction equipment production rate provides a comprehensive picture of work performance and helps in making appropriate decisions. The tower crane production rate is difficult to evaluate precisely since it fluctuates depending on the placement conditions, types of loads, and equipment allocation strategy. Moreover, when the production rate of tower cranes declines, the activity time and cost increase. Furthermore, as the time of the cycle lengthens, more fuel is consumed and more pollutants are produced, harming the environment (Golparvar-Fard et al., 2015).

In order to increase productivity, it is necessary to evaluate and monitor the tower crane's productivity throughout the execution phase in order to identify the equipment's inefficiencies and their underlying causes. This evaluation and monitoring are required to enhance productivity. To collect the data necessary for monitoring the performance of the equipment, however, requires a substantial investment of both time and resources (Chen et al., 2020). For instance, gathering data manually is not only fraught with the possibility of error but also impractical for large projects (J. Kim et al., 2019). This highlights the significance of automating the process of collecting data on equipment operation, analyzing its productivity, and monitoring its performance in large-scale building projects. Specifically, this emphasizes the need to automate the data collection process for equipment operation. Therefore, there is a research gap that needs to be covered in the scientific direction of measuring the actual tower crane production rate in construction sites. So, the purpose of this research is to develop a VRF with a user-friendly interface for a practical and speedy measure of tower crane actual cycle time on construction sites. The VRF will analyze the recorded videos captured by a video camera installed on the site to measure the tower crane's actual cycle time, which will be used in production rate calculations. Two case studies will be utilized to test and validate the VFR based on the cycle time.

2 LITERATURE REVIEW

2.1 Production Rate Monitoring

Typically, the project life cycle typically consists of five (5) distinct phases: initiation, planning, execution, monitoring, and closure. These phases work together to transform a project charter into a deliverable capable of carrying out its intended purpose (Radman et al., 2022). In terms of cost and time, the monitoring phase is responsible for coordinating various aspects. These factors consist of the phases of design, procurement, logistics, and installation (resources, materials, and equipment). The conventional method for monitoring productivity entails manually collecting data from the field, typically via meetings, and then converting that data into digital formats such as spreadsheets. Nevertheless, this activity is labor-intensive, has a low rate of productivity, and frequently results in delays and site job rework, which refers to the accuracy and speed with which data and information are delivered. As a result, project decision-makers are left in an uncertain state (Alizadehsalehi and Yitmen, 2021).

The definition of productivity varies based on the context in which it is discussed. The Oxford dictionary defines productivity as "the effectiveness of productive effort, especially in industry, as measured by the output per unit of input." Productivity indicates how much output can be generated for each unit of input. The unit rate is the number of actual work hours required to complete the specified quantity of construction work. This is the most prevalent method for calculating construction efficiency (Jang et al., 2011). Despite this, the units of measurement change based on the types of inputs and outputs involved in the construction process. Despite this, productivity's importance in reducing costs and generating profit is of the utmost importance in all industries, including the construction industry. Various technologies, including sensors, audio signals, and computer vision, were utilized by researchers in order to collect data on productivity.

A. Sensors

(Montaser and Moselhi, 2012) proposed a radio-frequency identification system for monitoring earth-moving operations (RFID). Their method was capable of automatically determining a truck's loading, traveling, unloading, and returning states. Because this strategy employs fixed RFID readers for gate systems at loading and dumping areas, it is best suited for use in projects with permanently installed loading and dumping areas. In addition, this approach cannot be used to calculate the length of time that vehicles wait in zones designated for loading and unloading. Separately (Montaser and Moselhi, 2014) created an automated system that integrates a global positioning system (GPS) with a geographic information system (GIS). This technique uses GPS gadgets installed on vehicles to monitor their whereabouts and uses a GIS to define the geographical limits of regions utilized for loading and dumping. Similar to their earlier technique, (Montaser and Moselhi, 2014) found the same four states for trucks; however, their system was unable to identify waiting periods at loading/dumping zones. This was an issue with their technique. This was because the system was never designed to function in this manner.

In order to address this limitation and improve the accuracy of measuring the volume of excavated soil, (Ibrahim and Moselhi, 2014) developed an automated method for evaluating the productivity of earth-moving operations by collecting data from five sensors mounted on the equipment. This was done to improve the accuracy with which the volume of excavated soil was measured. A number of parameters, including load queue, load type, travel, dump queue, and return, are maintained by the algorithm that was developed for recognizing the activities of equipment. The margin of error for the newly developed method for calculating productivity was only 2.2%. This method needs substantial installation of a large number of sensors on trucks and loaders, which is seldom feasible in construction projects owing to accessibility and availability constraints, as well as ownership models for heavy equipment

An accelerometer installed within the cabin of a medium-sized excavator was used to capture data at a frequency of one hundred hertz (Ahn et al., 2012). Utilizing vibration signals, they were able to establish the relationship between operational efficiency and environmental performance. (Ahn et al., 2015) conducted an experiment in a laboratory to collect acceleration data from four distinct excavator models. Each excavator's cabin contained an accelerometer that was used to gather these signals. In order to gather the necessary data for the experiment, which involved the analysis of accelerometer data patterns, the excavator had to be operated in accordance with very specific instructions. Various supervised classifiers, such as naive Bayes, instance-based learning, K-nearest neighbor (KNN), and decision trees, were employed to classify various characteristics of excavator operation (Ahn et al., 2015), which achieved more than 93% categorization accuracy. Methods for detecting the loading and unloading of a dumper truck employing remote tracking with three-axis magnetic field sensing and three-axis tilt sensing for both the loader and the vehicle were examined in a controlled laboratory setting. These sensing techniques were used on both the loader and truck.

(Akhavian and Behzadan, 2012) utilize remote tracking in combination with three-axis magnetic field sensing and three-axis tilt sensing. The authors utilize a 100 Hz GPS sensor, a three-axis accelerometer, and a three-axis gyroscope, to recognize equipment actions and the durations of such activities has been devised. Moreover, (Akhavian and Behzadan, 2015) proposed an approach for front-end loader using simulation input modeling. By utilizing a broad array of supervised learning algorithms, such as logistic regression, KNN, decision trees, neural networks, and support vector machines, this methodology was able to attain an overall accuracy of 86%. In some of the studies, smartphones' accelerometers and gyroscopes served as inertial measurement units (IMUs) to identify the various equipment types. The operation cycle time of an excavator was computed by (Kim et al., 2018) using IMU data collected at 128 Hz. By utilizing random forests, naive Bayes, decision trees, and sequential minimal optimization, they were able to predict the cycle time with a 91.83 percent degree of accuracy. In a separate study (Rashid and Louis, 2019), time series data augmentation was applied to data collected at 80 Hz from three-axis accelerometers and three-axis gyroscopes. The objective of this study was to generate training simulation data for four distinct models of excavators and front-end loaders. Utilizing a recurrent neural network (RNN), this technique was able to achieve an accuracy of better than 96% when applied to a fourfold augmentation.

(Kassem et al., 2021) developed a deep neural network (DNN) model for estimating the amount of soil dug by a mixed fleet of excavators (e.g., various sizes, weights, and models) and compared the excavation task performance using telematic data from 21 days of operation. The accuracy of 69.64% was deemed satisfactory because archaeologists were employed alongside their equipment in the case study chosen to represent central London. (Bae et al., 2019) utilize joystick signals, to develop a dynamic time-warping algorithm for activity recognition

and automatic classification of excavator activities such as digging, leveling, lifting, trenching, traveling, and idling. This algorithm was used to identify excavation, leveling, lifting, and trenching. The recognition accuracy of their model ranged between 91% and 97% of the time.

B. Audio Based Methods

Audio provides an additional data source for identifying the activities performed by heavy machinery due to the characteristic acoustic patterns that are typically produced by such machinery while it is performing its routine tasks (Cheng et al., 2019). Construction equipment activities were divided into two categories by (Cheng et al., 2017): productive activities, also known as major activities, and non-productive activities, also known as minor activities. This was utilized so that equipment states could be identified by listening to the sounds made by construction equipment. The research of (Sabillon et al., 2020), presents a model that uses audio data to estimate the cycle time of various types of machinery.

In comparison to computer vision techniques and sensors, audio signals are easier to use because they are captured by technologies such as microphones, which can cover a large area. This allows microphones to detect and record audio signals. Additionally, audio file processing requires fewer available computational resources (Sabillon et al., 2020). However, surrounding noise has the potential to impair the accuracy of the models, and the operation of certain machines does not produce identifiable sound patterns, making it impossible to detect the nature of the activity being performed by such machines (Cheng et al., 2017). This work contributes to this research area by designing and testing a novel technique for precisely forecasting the productivity of tower cranes. The application takes use of an inexpensive, simple-to-install system that is based on algorithms for tracking a single item. With the aid of a system based on algorithms for monitoring a single target, the device was designed and tested.

C. Vision-Based Methods

(Torres Calderon et al., 2021) could improve vision-based activity analysis techniques by developing a new method for training computer vision algorithms with data synthesized from three-dimensional (3D) kinematically configurable models. In addition to their visual qualities and geographical context, visual data may give information on the physical movements of equipment, which is the fundamental benefit of using computer vision to boost the productivity of equipment. This data may be used to improve the overall performance of the device (Kim and Chi, 2020). This technology has a variety of disadvantages and challenges, including its susceptibility to environmental elements such as occlusions, lighting, and illumination conditions; camera shaking caused by wind; and picture blurring caused by precipitation, snow, and fog. In contrast, if a more theoretical approach to analyzing this strategy was taken, it will offer a lot of beneficial characteristics (Gong and Caldas, 2011).

Researchers have devised a vision-based approach for assessing excavator productivity in a recent study (Chen et al., 2020). On the other hand, it was discovered that this method required a significant amount of computational power and had a number of drawbacks, including a dependence on lighting conditions, camera viewpoints, the number of objects in the scene, and the background's motion. In addition, they achieved an accuracy of 83 percent when measuring productivity and 94 percent when measuring idle time. In the study conducted by (Bügler et al., 2017), photogrammetry and video analysis were combined to determine the volume of soil removed during excavation and the efficiency of this process. By using computer vision and simulation, a model was created to analyze the productivity of earth-moving machinery by (H. Kim et al., 2019). The site access log was generated by identifying the license plates of dump trucks from surveillance camera footage captured at the entrance and exit of a construction site. Then, they utilized a simulation model to assess the productivity of trucks.

Vision-based approach applications for automating and monitoring production rates in the construction industry have only recently become popular, and they are still considered the novel. Furthermore, the current vision-based approach application result confirmed the optimal success in measuring and evaluating cycle time and production rate for construction equipment, according to the past literature review. However, there is a gap in measuring and evaluating the cycle time of tower cranes as they have had different characteristics from other equipment. Single object tracking algorithms and detection free tracking methods will be used in the proposed framework.

2.2 Single Object Tracking Algorithms

Tracking refers to the technique of identifying an item frame by frame inside a video. Tracking is a broad notion in computer vision and machine learning, despite the apparent simplicity of its meaning. It includes many unique concepts that are theoretically comparable but technically distinct. To monitor tower cranes, the OpenCV library,

which is an open-source, cross-platform computer vision and machine-learning software library, may be used with the appropriate tracking algorithms. OpenCV is a software library for computer vision and machine learning. While Single Object Tracking (SOT) (Brdjanin et al., 2020) is the level of tracking algorithms in OpenCV, which seeks to monitor a single object of a single class rather than many items. Visual object tracking is an alternate term for this process (Roberts and Golparvar-Fard, 2019). In SOT, tracked objects need to be defined by the user in the first frame by using the target object's bounding box as shown in Figure 1. Then the tracking algorithms will identify the tracked object in the remaining frames. SOT comes under the classification of detection free tracking since the first bounding box must be manually given to the tracker. This means that single item trackers must be able to monitor any object supplied to them, regardless of whether a categorization model has been established for that object (Bradski, 2000). The eight single object tracking algorithm that available in OpenCV are:



Figure 1: An example of bounding box that detect the boom of a tower crane

BOOSTING, often known as online boosting (Grabner et al., 2006), is the online edition of AdaBoost, from whence it derives its name. It operates based on the categorization of the monitored item and the backdrop and is trained in real-time. Consequently, it begins with a bounding box input of the item (positive class), either manually or through an object detector, and trains itself against patches created around the bounding box (negative background class). For each new frame, the classifier is executed in the region of the previous position, and the place with the highest classification score is chosen as the new object position, while the classifier is also updated with the freshly gathered data.

The Multiple-Instance-Learning (MIL) tracker was formerly referred to as MILBoost. As the name suggests, it employs a similar methodology to the previous tracker idea; however, instead of assessing a part of the processed frame as a whole, it divides it into sub-sections that are then separately categorized and summarized as a collection of samples in so-called bags (Babenko et al., 2009). These bags are then categorized as positive (containing an item) or negative (not containing an object) depending on the findings of their collected samples, without needing all of their samples to be categorized as positive (in fact one positive sample in a bag can be enough). By using this method, the tracker has a more precise understanding of what a positive class may be like when occlusion or centering issues of the whole object occur. While it might enhance accuracy, it is anticipated to operate more slowly.

In response, a quicker on-line change (Henriques et al., 2015) was suggested, which is an abbreviation for Kernelized Correlation Filters (KCF) (Shin et al., 2020), builds on BOOSTING and MIL, but mathematically uses the natural overlapping of samples in the MIL method and the extra data that comes with it to provide a substantially quicker and more accurate tracking system. As its name suggests, it uses the Fourier Transform to execute a computationally efficient correlation calculation that is responsible for the significant acceleration. In addition to speed and accuracy benefits, it is much better to detect a tracking failure rather than continuing to track a different object after a tracking loss.

MedianFlow operates by following individual points of a monitored object both forwards and backwards in successive frames and comparing their trajectories, which should overlap if the tracking is ideal (Kalal et al., 2010). Any observed disparities between the trajectories are then utilized to assign tracking faults to locations inside the bounding box of the tracked item and to classify these points as in- and outliers. The filtering of outliers and the prediction of the bounding box's motion based on the inliers make the tracker particularly effective for predicting movements. In addition, this method enables the tracker to quickly identify a failed tracking operation so that it

does not follow the incorrect objects. On the downside, MedianFlow is more sensitive to occlusion and large or unstable target object movements.

Tracking-Learning-Detection (TLD) was promoted by the author (Kalal et al., 2012) as a tracking framework rather than a simple tracking algorithm. The framework is to divide the detection, tracking, and learning functions into distinct components to improve the overall tracking performance over time. One of the primary goals of the idea is to prevent on-line learning in cases when the item to be monitored is no longer in the tracking frame (e.g., due to conclusion), since this leads to negative learning. In contrast, the detector component must more precisely locate observed target appearances in order to correct the tracker in the event of divergence, whilst the learning component must analyze and estimate the mistakes produced by the detector and update it to minimize them. As a result, the tracking framework must be able to re-detect even lost target objects when they reappear on the input frame, but the method may also result in unintended hops between the target item and other and/or similar objects in its vicinity.

The Generic Object Tracking Using Regression Networks (GOTURN) tracker It is the only tracking algorithm throughout the tracker class that uses a convolutional neural network (CNN). This technique discovers a general association between object motion and appearance and may be used to monitor items not included in the training set. Online trained tracker algorithms are sluggish and do not function effectively in real time since they cannot utilize a huge number of videos to increase performance. In contrast, offline tracker algorithms may be trained to handle rotations, perspective shifts, illumination variations, and other difficult issues. However, GOTURN use a regression-based method to track objects with a single feed-forward run across the network, they essentially regress straight to find target objects. Two inputs are received by the network: a search area from the current frame and a target from the previous frame. The network then compares these photos to the current image to locate the target item (Held et al., 2016).

The Minimum Output Sum of Squared Error (MOSSE) tracker also works with correlations in the fourier domain, but it does so considerably more quickly than similar trackers while still producing excellent tracking outcomes. The method of minimizing a Sum of Squared Error (SSE) is not novel, but that previous applications typically assumed the object of interest was located in the center of a frame, which was not always the case during tracking operations (Bolme et al., 2010). Initialization is followed by tracking as the two primary components of the MOSSE tracking algorithm. Initialization may be accomplished in one of two ways: with a collection of initialization frames to establish a tracking filter, or with a single start frame on which the filter is constructed. Using just one initialization frame is the more difficult form, and it is the one used by OpenCV's MOSSE implementation. This sophisticated version also continually updates the tracking filter, unlike the simpler version with many startup frames. Consequently, MOSSE becomes less sensitive to differences in size, position, deformation, and illumination. It relies on the tracker's learning rate, or how quickly it can adjust to changing tracking circumstances.

The Discriminative Correlation Filter with Channel and Spatial Reliability (CSRT)'s inclusion of both spatial and channel dependability makes the Discriminative Correlation Filter (DCF) algorithm more successful (Lukezic et al., 2017). The CSRT tracker is better to the normal DCF algorithm due to its capacity to adapt non-rectangular targets in line with the spatial reliability map. In practice, the spatial reliability map is used to fine-tune filter support for the specific tracking zone that has been selected. Each channel's dependability is evaluated, and the findings are merged to create the final response map. In this manner, the significance of each channel filter is determined. Using simply the HOG and color histograms, the CSRT tracker is able to achieve high object tracking precision in a very short period of time.

According to (Brdjanin et al., 2020; Haggui et al., 2021), the BOOSTING tracker is solely used for legacy purposes and as a comparison tool for other algorithms, despite being slower and less effective than other trackers. The MIL tracker offers much better degrees of precision than the BOOSTING tracker. However, the most deficiency of the MIL tracker is that it contains reporting mistakes. The KCF tracker is substantially more efficient than BOOSTING and MIL. The TLD tracker is plagued by a considerable number of false positives, which has led to a variety of implementation difficulties. The Median Flow performs an excellent job of reporting failures; however, this model will fail if it encounters an abrupt change in appearance, a huge leap in motion, or quick motion.

Compared to CSRT and KCF, MOSSE's accuracy is worse at the price of its outstanding speed. This tracker is a good alternative if concentrate entirely on speed is needed. Despite being significantly slower, the CSRT tracker

is comparable to the KCF tracker in terms of accuracy. Consequently, KCF trackers are versatile enough to be used in nearly any scenario within the construction industry. However, due to the dynamic nature of the construction industry and the unpredictability of the circumstances and events that may occur during the execution phase, the other seven algorithms will be integrated in the proposed framework to overcome any potential eventualities.

The bottom line is that each of the previously mentioned eight trackers has some unique features and capabilities which are considered optimal for the construction site nature and circumstances. Therefore, to benefit from all these features and capabilities, this research will utilize the eighth trackers to develop a VRF with a user-friendly interface. Furthermore, the VRF performance and accuracy will be enhanced and confirmed as it will be suitable for all construction site circumstances. This VRF will be used for a practical and speedy measure of tower crane actual cycle time on construction sites. The VRF will analyze the recorded videos captured by a video camera installed on each site while enabling the user to easily choose the suitable tracker for each site. However, the VRF working method will be explained in detail in the following sections.

3 METHODOLOGY

This research proposes a VRF for measuring and calculating tower cranes' actual cycle times on construction sites. The VRF will analyze the recorded videos captured by a video camera on the site by a sophisticated computer program to measure the tower crane's actual cycle time that will be used in tower crane production rate calculations. The video camera is utilized to capture the tower crane in full motion during material loading, movement while laden, material unloading, and movement when empty to return to its loading position. Although the tower crane operates mainly in appropriate weather conditions according to safety precautions. However, the number of videos (more than four videos) must be captured in anticipation of any bad weather conditions (fog or rain) that may occur and affect the quality of the recorded videos. The VRF will efficiently track and analysis the cycle time of tower crane from videos.

The VRF was developed by using the OpenCV library in the Python programming language due to analyzes of the captured videos. This code will enable the VRF to monitor and detect the boom of the tower crane movement from captured videos. The algorithm of the VRF is shown in Algorithm 1. The videos will be divided into frames, and the VRF will keep track of the boom throughout each frame of captured videos to record a history of where that boom was over time to collect more precise data. By the end of the analysis process, it will be possible to estimate and measure the tower crane cycle time and the number of cycles completed during the video time. Consequently, it will be possible to calculate the tower crane production rate through PR calculation equations. Additionally, the captured videos can serve as a clear and transparent evidence basis for monitoring and management of the tower crane activities if some of them take a longer time than typically estimated or measured.

Algorithm 1

Input: Video address and tracking algorithm choice (BOOSTING, MIL, KCF, TLD, MEDIANFLOW, GOTURN, MOSSE, or CSRT)

Output: cycle time, fps, status, and count of movements

Initialization:

Select the tower crane by the user

Select the region to monitor by the user

Fps = none, frame skip = 0

For all frames of the video **do**

If tower crane detected **then**

 Draw the bounding box of the tower crane and the wanted region to be monitored

 Count the cycle time, fps, status, and movements

Else

Frame skip

return updated *fps*, the cycle time, the count of movements

The VRF interface is shown in Figure 2 part (a). Part (b) shows the tracker number, which will be an input in the “Tracker Number” cell. Therefore, to run the VRF, first the user selects both the video and the tracking algorithm through the software application interface. Second, the user chooses the suitable tracking algorithm for his needs as shown in part (c). After that, the user runs the video as shown in part (d). A new interface will be available as shown in Figure 3. In which the user will define the region of interest for the boom of the tower crane as shown in part (a), and then locate the region where the boom must bypass to continue its movement as shown in part (b). After that, the algorithm tracks the boom of the tower crane through one of the selected trackers and displays a bounding box around it during its movement. The tracking process continues till the end of the video, after which the program will run and count cycles and time as shown in part (c). Also, to close the video, the user can press the “Q button” on the keyboard.

Measuring the cycle time of a tower crane depends on four main objectives (loading time, moving time, unloading time, and return time). For all these, the program will calculate it and the output time will be the cycle time that it took the tower crane to transport the package. Also, evaluating the production rate of tower cranes, which involves numerous variables, can be a difficult operation, and it can be described as a measurement of the machine’s output (i.e., kg/hr). (Šopić et al., 2021) provide a comparable formula (Eq. 1) for the production rate, where the production rate measure (PR) is the ratio of production quantity (PQ) and time spent on production (TP). The mentioned formula (Eq. 1) and unit measure (kg/hr) are proposed to be used after calculating the cycle time by the (VRF).

$$PR = \frac{PQ}{TP} \quad \text{Eq. 1}$$

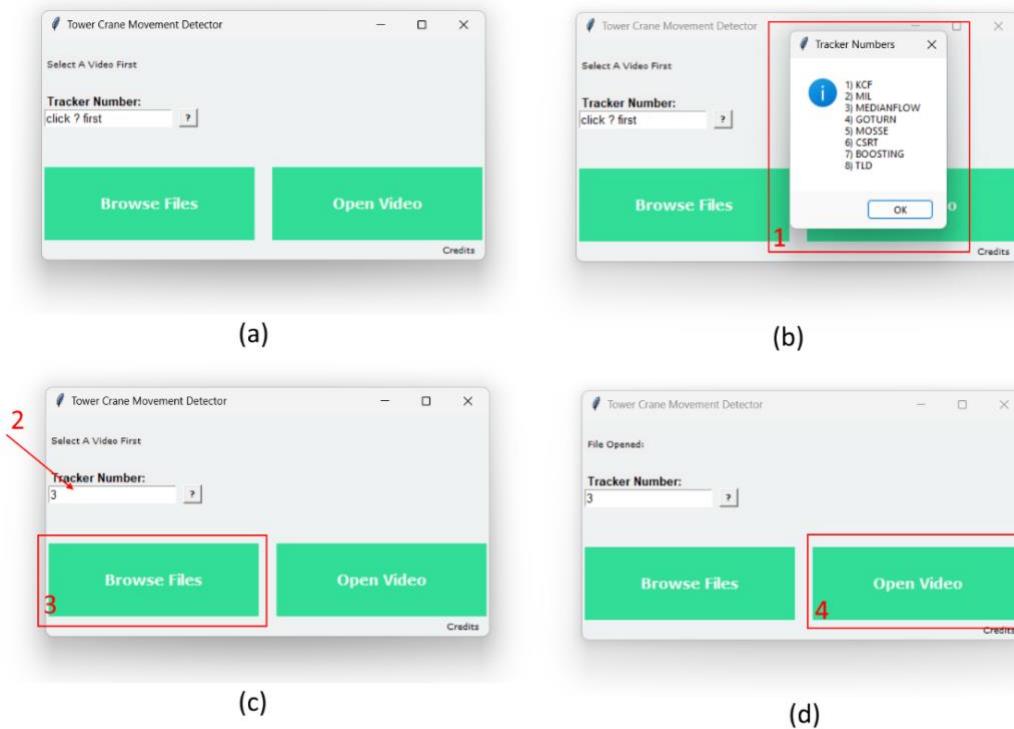


Figure 2: The VRF interface

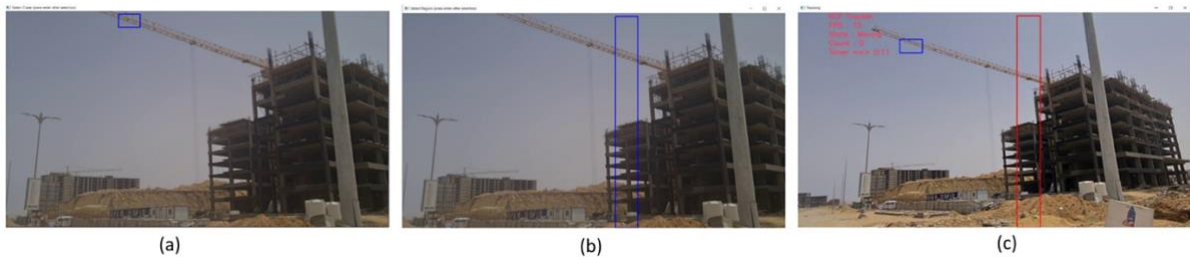


Figure 3: The steps of assigning the tower crane and the region to monitor

4 CASE STUDIES

4.1 Description of the Case Studies

The first case study was a construction site located in Kafr Al-Sheikh, Egypt, where tower crane operations were performed at the time of recording. Five short videos of tower cranes were recorded during work. At the observed construction site, twenty-nine tower cranes were used for the construction work. One crane was then selected to be focused on. This crane was in good shape and well maintained. The crane was 51 meters in height and could lift up to a maximum load of 8 tons, and the trolley had a range of 65 meters. Moreover, the second case study was on a construction site located in the administrative capital of Egypt. The project is a construction consisting of two bedrooms, ground floor, and eight floors, which is a small project. This construction was served by one tower crane of the type “Young Mao stt139.” This tower crane can lift a load of up to six tons at a height of twenty meters. The tower crane was built 48 meters high, the trolley was 60 meters long, and the distance the trolley could move was 55 meters. A conclusion for the tower crane specifications for both case studies is shown in Table 1.

Table 1: The tower cranes specifications of the both case studies

Term	Case study (1)	Case study (2)
Number of tower cranes	1	1
Lift up load (ton)	8	6
Hight (meters)	51	48
Boom length (meters)	65	60

4.2 Data Collection and Analysis of the Case Studies

For the first case study, five videos were captured during the execution of the project. A manual calculation for each video was calculated by a stop watch. The first video after using the stop watch resulted in 256 sec. Moreover, the other four videos were calculated with the same concept. Only one of the two cranes was used in the calculation. On the other hand, the automated tool which is used by the proposed model VRF, KCF algorithm, was used in the five captured videos. The calculation of the cycle time was calculated for each video by the VRF. The results of the manual calculation and the automated calculation are shown in Table 2, where original time means the time resulting from the use of the stopwatch, and estimation time means the time produced after the videos are entered into the program to analyze and determine the time of one cycle. The sensor’s rated load was 2,000 kg, and the production rate was 14,458 kg/hr. when Eq. 1 was used by hand and 13,900 kg/hr. when the VRF was used.

Table 2: The compression between the manual method and the tool to detect the cycle time of five videos for each case study

Term	Case Study (1)		Case Study (2)	
	Original	Estimation	Original	Estimation
Cycle time video 1	256 sec	232 sec	255	246
Cycle time video 2	422 sec	411 sec	264	209
Cycle time video 3	568 sec	731 sec	252	240
Cycle time video 4	387 sec	380 sec	254	247
Cycle time video 5	859 sec	836 sec	253	221
Avg. cycle time	498 sec	518 sec	255 sec	232 sec
Sum cycle time	2492 sec	2590 sec	1278	1163
Production rate	14,458 kg/hr.	13,900 kg/hr.	7,042 kg/hr.	7,739 kg/hr.

The second case study included the recording of an additional five videos documenting the execution of the work during the day. In both the manual calculation and the automated calculation produced by the proposed VRF, the same underlying concepts were used. Table 2 displays the results of both the manual and computerized calculations that were performed. The sensor had a maximum recommended load capacity of five hundred kilograms. The production rate was determined by dividing the total amount transferred by the typical cycle time; the production rate that was obtained through manual calculation was 7042 and the automated calculation was 7739 kg/hr. In addition, to calculate the accuracy of the proposed VRF, two evaluation methods were adopted, which are the mean absolute percentage error (MAPE) as shown in Eq. 2 and MAPE in Eq. 3. Based on calculations, it appeared that the result of the accuracy of the program ranged from ninety percent to ninety-five percent. The results for the evaluation of the model are shown in Table 3.

$$APE = \frac{|Actual - Forecast|}{|Actual|} \quad \text{Eq. 2}$$

$$MAPE = \frac{1}{Sample\ Size} * \sum_{t=1}^n APE * 100 \quad \text{Eq. 3}$$

Table 3 - The absolute percentage error (APE), MAPE, and the Median APE for the ten detected videos

term	Case Study (1)	Case Study (2)
Cycle time video 1 APE	9,38	3.53
Cycle time video 2 APE	2.61	20.83
Cycle time video 3 APE	28.70	4.76
Cycle time video 4 APE	1.81	2.76
Cycle time video 5 APE	2.68	12.65
MAPE	9.03	8.91
Accuracy of MAPE	90.97%	90.27%
Median APE	2.61	3.53
Accuracy of median APE	97.32%	95.24%

5 CONCLUSION

Time and resources are required to collect the information necessary for monitoring the performance of the equipment, especially tower cranes. In large construction projects, it is necessary to automate the process of collecting data on tower crane activities, evaluating, and analyzing their productivity, and monitoring their performance. Therefore, the VRF for automating the calculation of cycle time was developed based on the OpenCV library, which adopts eight algorithms to track and monitor the tower crane's cycle time. The VRF was developed by adopting eight single tracking detection free algorithms. To evaluate the VRF, two case studies were used to evaluate the proposed framework in Egypt; Kafr El-Sheikh and the administrative capital. Five captured videos for each case study to evaluate the model. The manual method was conducted by using a stop watch for each video to calculate the cycle time for the case study's tower crane. Then the proposed VRF was used to calculate the cycle time of the tower crane automatically. After analysis of the videos and calculate the production rates of the two case studies, the proposed model was achieved accuracy 97% and 95%, respectively.

ACKNOWLEDGMENTS

The authors would like to thank everyone who helped create the video from the construction site.

REFERENCES

- Ahn, C.R., Lee, S., Peña-Mora, F., 2015. Application of low-cost accelerometers for measuring the operational efficiency of a construction equipment fleet. *Journal of Computing in Civil Engineering* 29, 04014042.
- Ahn, C.R., Lee, S., Peña-Mora, F., 2012. Monitoring system for operational efficiency and environmental performance of construction operations using vibration signal analysis, in: *Construction Research Congress 2012: Construction Challenges in a Flat World*. pp. 1879–1888.
- Akhavian, R., Behzadan, A.H., 2015. Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers. *Advanced Engineering Informatics* 29, 867–877.
- Akhavian, R., Behzadan, A.H., 2012. Remote monitoring of dynamic construction processes using automated equipment tracking, in: *Construction Research Congress 2012: Construction Challenges in a Flat World*. pp. 1360–1369.
- Alaghbari, W., Al-Sakkaf, A.A., Sultan, B., 2019. Factors affecting construction labour productivity in Yemen. *International Journal of Construction Management* 19, 79–91.
- Alizadehsalehi, S., Yitmen, I., 2021. Digital twin-based progress monitoring management model through reality capture to extended reality technologies (DRX). *Smart and Sustainable Built Environment*.
- Babenko, B., Yang, M.-H., Belongie, S., 2009. Visual tracking with online multiple instance learning, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 983–990.
- Bae, J., Kim, K., Hong, D., 2019. Automatic identification of excavator activities using joystick signals. *International Journal of Precision Engineering and Manufacturing* 20, 2101–2107.
- Bolme, D.S., Beveridge, J.R., Draper, B.A., Lui, Y.M., 2010. Visual object tracking using adaptive correlation filters, in: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 2544–2550.
- Bradski, G., 2000. The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 25, 120–123.
- Brdjanin, A., Dardagan, N., Dzigal, D., Akagic, A., 2020. Single object trackers in opencv: A benchmark, in: *2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*. IEEE, pp. 1–6.
- Bügler, M., Borrmann, A., Ogunmakin, G., Vela, P.A., 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., Zhu, Z., Hammad, A., 2020. Automated excavators activity recognition and productivity analysis from construction site surveillance videos. *Automation in construction* 110, 103045.
- Cheng, C.-F., Rashidi, A., Davenport, M.A., Anderson, D.V., 2019. Evaluation of software and hardware settings for audio-based analysis of construction operations. *International Journal of Civil Engineering* 17, 1469–1480.
- Cheng, C.-F., Rashidi, A., Davenport, M.A., Anderson, D.V., 2017. Activity analysis of construction equipment using audio signals and support vector machines. *Automation in Construction* 81, 240–253.
- Golparvar-Fard, M., Peña-Mora, F., Savarese, S., 2015. Automated progress monitoring using unordered daily construction photographs and IFC-based building information models. *Journal of Computing in Civil Engineering* 29. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000205](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000205)
- Gong, J., Caldas, C.H., 2011. An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations. *Automation in Construction* 20, 1211–1226.
- Grabner, H., Grabner, M., Bischof, H., 2006. Real-time tracking via on-line boosting., in: *Bmvc*. Citeseer, p. 6.
- Haggui, O., Tchalim, M.A., Magnier, B., 2021. A comparison of opencv algorithms for human tracking with a moving perspective camera, in: *2021 9th European Workshop on Visual Information Processing (EUVIP)*. IEEE, pp. 1–6.

- Held, D., Thrun, S., Savarese, S., 2016. Learning to track at 100 fps with deep regression networks, in: European Conference on Computer Vision. Springer, pp. 749–765.
- Henriques, J.F., Caseiro, R., Martins, P., Batista, J., 2015. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 583–596. <https://doi.org/10.1109/TPAMI.2014.2345390>
- Ibrahim, M., Moselhi, O., 2014. Automated productivity assessment of earthmoving operations. *Journal of Information Technology in Construction (ITcon)* 19, 169–184.
- Jang, H., Kim, K., Kim, J., Juhung, Kim, Jaejun, 2011. Labour productivity model for reinforced concrete construction projects. *Construction Innovation*.
- Kalal, Z., Mikolajczyk, K., Matas, J., 2012. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 1409–1422. <https://doi.org/10.1109/TPAMI.2011.239>
- Kalal, Z., Mikolajczyk, K., Matas, J., 2010. Forward-backward error: Automatic detection of tracking failures, in: 2010 20th International Conference on Pattern Recognition. IEEE, pp. 2756–2759.
- Kassem, M., Mahamedi, E., Rogage, K., Duffy, K., Huntingdon, J., 2021. Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach. *Automation in Construction* 124, 103532.
- Kim, H., Ahn, C.R., Engelhaupt, D., Lee, S., 2018. Application of dynamic time warping to the recognition of mixed equipment activities in cycle time measurement. *Automation in Construction* 87, 225–234.
- Kim, H., Ham, Y., Kim, W., Park, S., Kim, H., 2019. Vision-based nonintrusive context documentation for earthmoving productivity simulation. *Automation in Construction* 102, 135–147. <https://doi.org/10.1016/j.autcon.2019.02.006>
- Kim, J., Chi, S., 2020. Multi-camera vision-based productivity monitoring of earthmoving operations. *Automation in Construction* 112. <https://doi.org/10.1016/j.autcon.2020.103121>
- Kim, J., Chi, S., Choi, M., 2019. Sequential pattern learning of visual features and operation cycles for vision-based action recognition of earthmoving excavators, in: *Computing in Civil Engineering 2019: Data, Sensing, and Analytics*. American Society of Civil Engineers Reston, VA, pp. 298–304.
- Kim, J.Y., Caldas, C.H., 2013. Vision-based action recognition in the internal construction site using interactions between worker actions and construction objects. Presented at the ISARC 2013 - 30th International Symposium on Automation and Robotics in Construction and Mining, Held in Conjunction with the 23rd World Mining Congress, pp. 661–668. <https://doi.org/10.22260/isarc2013/0072>
- Lukezic, A., Vojir, T., Čehovin Zajc, L., Matas, J., Kristan, M., 2017. Discriminative correlation filter with channel and spatial reliability, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 6309–6318.
- Montaser, A., Moselhi, O., 2014. Truck+ for earthmoving operations. *Journal of Information Technology in Construction (ITcon)* 19, 412–433.
- Montaser, A., Moselhi, O., 2012. RFID+ for tracking earthmoving operations, in: *Construction Research Congress*. pp. 1011–1020.
- Radman, K., Jelodar, M.B., Lovreglio, R., Wilkinson, S., Ghazizadeh, E., 2022. Real-Time Project Productivity Tracking System: Practical Case in Smart Construction Projects, in: *7TH NEW ZEALAND BUILT ENVIRONMENT RESEARCH SYMPOSIUM*. p. 409.
- Rashid, K.M., Louis, J., 2019. Times-series data augmentation and deep learning for construction equipment activity recognition. *Advanced Engineering Informatics* 42, 100944.
- Roberts, D., Golparvar-Fard, M., 2019. End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level. *Automation in Construction* 105. <https://doi.org/10.1016/j.autcon.2019.04.006>

- Sabillon, C., Rashidi, A., Samanta, B., Davenport, M.A., Anderson, D.V., 2020. Audio-based bayesian model for productivity estimation of cyclic construction activities. *Journal of computing in civil engineering* 34, 04019048.
- Shin, J., Kim, H., Kim, D., Paik, J., 2020. Fast and robust object tracking using tracking failure detection in kernelized correlation filter. *Applied Sciences* 10, 713.
- Šopić, M., Vukomanović, M., Car-Pušić, D., Završki, I., 2021. Estimation of the excavator actual productivity at the construction site using video analysis. *Organization, Technology and Management in Construction: an International Journal* 13, 2341–2352. <https://doi.org/10.2478/otmcj-2021-0003>
- Torres Calderon, W., Roberts, D., Golparvar-Fard, M., 2021. Synthesizing Pose Sequences from 3D Assets for Vision-Based Activity Analysis. *Journal of Computing in Civil Engineering* 35. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000937](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000937)