Fusion of Photogrammetry and Video Analysis for Productivity Assessment of Earthwork Processes

M. Bügler[1], G. Ogunmakin[2], P. A. Vela[2], A. Borrmann[1], J. Teizer[3]

[1]Chair of Computational Modeling and Simulation, Technische Universität München, Arcisstraße 21, 80333 Munich, Germany
[2]School of Electrical and Computer Engineering, Georgia Institute of Technology, United States of America
[3]RAPIDS Construction Safety and Technology Laboratory, Ettlingen, Germany

Abstract: The high complexity of modern large scale construction projects leads their schedules to be sensitive to delays. At underground construction sites the earthwork processes are vital, as most of the following tasks depend on them. This paper presents a method for estimating the productivity of soil removal by combining two technologies based on computer vision: photogrammetry and video analysis. Photogrammetry is applied to create a time series of point clouds throughout excavation, which are used to measure the volume of the excavated soil for daily estimates of productivity. Video analysis is used to generate statistics regarding the construction activities for estimating productivity at finer time scales, when combined with the output from the photogrammetry pipeline. As there may be multiple causes for specific productivity levels, the automated generation of progress and activity statistics from both measurement methods supports interpretation of the productivity estimates. Comparison to annotated ground truth for the tracking and activity monitoring method highlights the reliability of the extracted information. The suitability of the approach is demonstrated by two case studies of real-world urban excavation projects.

1 INTRODUCTION

Construction sites involve significant quantities of resources, including multiple types of manpower, equipment, and materials. Proper coordination of these temporary entities positively impacts on-site productivity, which in turn influences construction safety, costs, and schedules (Goodrum et al., 2009; Zhai & Goodrum, 2009). In a related vein, awareness of labor productivity has been shown to improve the direct work rate (Gouett & Haas, 2011). Thus, the existence of on-site productivity measurements is sufficient to improve on-site operations and mitigate any adverse conditions that may impede progress.

Progress tracking and resource utilization tracking constitute two distinct components of productivity measurement. Specifically, progress tracking measures quantities installed while resource utilization tracking measures consumed work hours as well as how such work hours were spent (Zhai & Goodrum, 2009). The key difference between the components lies in the definition of productivity. A resource can be fully utilized but not achieve any progress. An excavator can move soil within an excavation area all day, but not be productive in terms of the volume of soil removed from the excavation area within a certain time. Current techniques for site operation analysis, as described by Goodrum et al. (2009), focus on the monitoring of construction progress and the measurement of work task productivity, but are heavily based on manual efforts or at best partially automated. In real operations they are frequently out-of-date (Gong & Caldas, 2011).

Many current research efforts associated to progress and productivity monitoring seek to prove the hypothesis that it is possible to reliably track multiple resources with images (video and/or time-lapse) in order to reproduce the daily workflow activities associated to a construction site. The intent behind such monitoring and analysis is to automatically provide critical information, through computer-vision algorithms, on construction operations for improved decision making in construction engineering and management (Teizer et al., 2005).

This paper focuses on a vision-based approach to the automatic estimation of productivity associated to excavation processes on a construction site. Earthwork processes are often subject to unanticipated delays (Assaf & Al-Hejji, 2006), which are likely to propagate through the entire remaining schedule and adversely impact progress, productivity, and costs (Koehn et al., 1978). The presented work impacts research into excavation operations by enabling the automated monitoring and tracking of on-site resources. Video-based monitoring, photogrammetry and processing algorithms provide a non-intrusive, easy, inexpensive, and rapid mechanism for generating a body of operational information and knowledge on the progress and productivity of excavation operations. If made available to project stakeholders, the information and knowledge would enable inquiry into construction operations that is currently not possible (Bohn & Teizer, 2010). Longer term, vision-based research can serve as a valuable aid to project management by enabling tighter control and greater efficiency.

Demonstrating that a visual measurement system can effectively analyze and assess work-site productivity will assist project managers by reducing the time spent monitoring and interpreting project status and performance, thus enabling increased attention to the control of cost and schedule. By making project management and workforce more aware of the performance status of their project and their work environment, potential savings to the industry are envisioned. Since benefits in construction often impact a broader theme of issues, they are likely to impact schedule, cost, safety, and quality at the same time.

Contribution. This paper introduces a novel methodology which combines two different sources of data (photogrammetry and video analysis) to gain insight into the productivity of earthwork operations. Here, earthwork operations refers to the coordination between excavators and dump trucks for soil removal. The proposed work flow, illustrated in Figure 1, retrieves the excavated volume from point clouds created through photogrammetry algorithms, generates activity statistics of interacting excavators and dump trucks from automatically processed surveillance video, then fuses the information to arrive at soil removal productivity estimates over the course of a day (e.g., at hourly intervals). Statistical analysis of the excavated volume progress and the machine activity states provide supporting information for understanding the factors influencing the estimated productivity levels. By reacting upon this knowledge, a more optimal allocation of resources can be found, saving time and costs for project stakeholders.

2 RELATED WORK

While current research into construction site monitoring has considered non-vision technological solutions such as Radio Frequency Identification (RFID) and Ultra Wideband technology (Cheng et al., 2013; Costin et al., 2012; Grau et al., 2009; Saidi et al., 2011), and Global Positioning Systems (Grau et al., 2009; Pradhananga & Teizer, 2013; Vahdatikhaki & Hammad, 2014; Vasenev et al., 2014), the current focus of the related work is on vision-based solutions. Further, the emphasis is on passive imaging as opposed to active imaging such as that provided by laser scanners (Bosché, 2010). Although laser-scanning methods are known to have better accuracy and measurement density, passive imaging solutions have been shown to provide complementary capabilities that make them suitable to certain monitoring tasks (Golparvar-Fard et al., 2011; Malinovskiy et al., 2009). The review will first cover vision-based progress monitoring via 3D reconstruction of the construction site, and then cover vision-based activity estimation for on-site resources.

2.1 Image-Based Progress Monitoring.

The use of imagery to estimate construction progress and connect it to building information models (BIM) has been demonstrated to work using fixed cameras mounted around the construction site (Karsch et al., 2014; Lukins et al., 2007). Due to the geometric complexity and potential occlusions on larger work-sites, hand-held imaging devices are often preferable (Marzouk & Zaher, 2015). Successful methods use hand-held devices and photogrammetry (Braun et al., 2015; Dimitrov & Golparvar-Fard, 2014; Golparvar-Fard et al., 2011; Walsh et al., 2013), with some methods specialized to particular elements of the built infrastructure such as façades and roofs (Aydin, 2014; Fathi & Brilakis, 2013). Photogrammetry uses photos typically taken by a pedestrian worker equipped with a conventional digital camera as input data and generates, through computational algorithms, a three-dimensional point cloud of the imaged site. Contemporary research has also focused on unmanned platforms for performing the data collection (Mills et al., 2010; Siebert & Teizer, 2014; Zhang & Elaksher, 2012). The focus of this paper is on excavation processes and volumetric analysis. Preliminary work presented by Bügler et al. (2013) estimated the current progress of an excavation process by determining the excavated volume through the Visual Structure from Motion (VisualSFM) algorithm (Wu, 2011).

2.2 Vision-Based Resource Tracking

Using passive imaging cameras aimed at a worksite for performing resource tracking and activity monitoring relies on applying methods and tools from surveillance research (Collins et al., 2000). As will be detailed further in Section 4, video surveillance systems require the connection of several modules, which perform object detection, identification, tracking, and reidentification (needed when an object leaves and returns the scene). Once these basic components are functional, an additional interpretation module may be added in order to identify behaviors or activities engaged by the tracked objects within the sensed scene (Haritaoglu et al., 2000). These may be further decomposed into typical or unusual (Javed et al., 2003). Further analysis of the behaviors and activities over time may be performed in order to identify key events, which can then be entered into a database or spreadsheet for reporting or query purposes. For example, the system by Tian et al. (2008) performs event-based retrieval, provides real-time event alerts through the internet, and extracts long term statistical patterns of activity in a more general setting.

Detection. Resource detection on construction site videos is broadly categorizable into two approaches: object detection algorithms and general foreground esti-

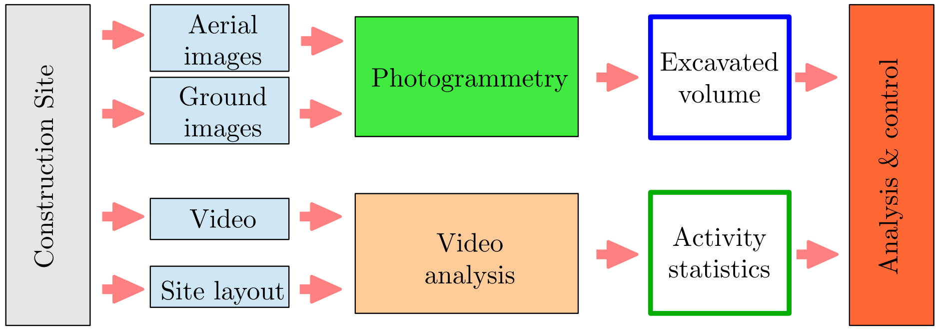


Figure 1: Data flow diagram of proposed concept

mation algorithms. Detection algorithms, usually relying on machine learning techniques, involve training to learn the unique signature of a given object. Algorithms include neural networks (Rezazadeh Azar & McCabe, 2012a), support vector machines with specific feature models (Memarzadeh et al., 2012), random forests (Park et al., 2011a), and parts-based models (Rezazadeh Azar & McCabe, 2012b). Parts-based modeling approaches work best for articulated objects since their appearance geometry has high variation, which can be compensated through multiple, individual part detectors. Since detection-based methods often seek specific targets, the resource type will typically be known from the detection itself. Foreground estimation algorithms tend to be simpler; they generate a model of the expected scene and classify continuous regions that do not match the model to be target regions (Chi & Caldas, 2011; Gong & Caldas, 2009). Foreground estimation works well when the entrance into the scene of the object is controlled, or the object of interest is not occluded by other foreground objects (Yang et al., 2011). In less controlled settings, stronger results are obtained by combining the two techniques (Chi & Caldas, 2011; Gong & Caldas, 2011; Rezazadeh Azar & McCabe, 2012a). The value is due to the fact that foreground detection methods do not classify the detected objects. Therefore, the addition of object recognition or object detection algorithms to classify the foreground regions has the added benefit of both rejecting irrelevant foreground detections and recognizing the resource category.

Tracking. When seeking to estimate target trajectories from target detections, additional processing is required to keep track of the detected objects over time. Situations such as short term occlusion, trajectory crossings, and object overlap, all lead to target label ambiguity and require data association methods to properly link detections across frames. A common approach is to utilize spatio-temporal reasoning to generate the proper associations (Chi & Caldas, 2011). An alternative approach is to incorporate a target tracking algorithm, which utilizes the detection region to generate a specific signature model for the target, then search for the signature in future frames independently from the detection results. There have been comparative papers studying the performance of specific tracking algorithms on construction worksites (Park et al., 2011b; Teizer & Vela, 2009) with probabilistic and kernel-based methods showing strong performance. Kernel-based methods have been used since for tracking on-site resources (Park & Brilakis, 2012). Yang et al. (2010) used a probabilistic kernel method to track multiple workers with overlapping trajectories, showing that these methods can be modified to handle occlusions. Ogunmakin et al. (2013) extended these results to rigid construction machines observed from a distance. While most tracking papers are on tracking with a single camera view, Brilakis et al. (2011) demonstrated 3D tracking of construction site resources using stereo video.

Activity Analysis. For a detected and tracked construction resource, further analysis of the object's visual stream provides important information regarding the role and contribution of the resource to the construction process. Deciphering this information falls within the category of activity analysis. Activity analysis on a construction site involves determining the action each target in the scene is engaged in over a period of time. Early activity analysis utilized sensors installed on the resource of interest, with demonstrated proof of concept (Ake, 2001). Since then, research has explored the feasibility of vision-based strategies by translating advances in computer vision to the construction field. The vision-based research literature, not just in construction but in general, can be split into activity identification through analysis of specific spatio-temporal visual features of the resource or through analysis of the trajectory and proximity information.

The former category has mostly focused on articulated resources, such as personnel and machines (Peddi et al., 2009; Weerasinghe & Ruwanpura, 2010). By decoding the articulation poses or target feature elements over time, the activity category can be inferred (Khosrowpour et al., 2014; Yang et al., 2015). Work activities may be broken into effective work, ineffective work, or contributory work for productivity analysis (Peddi et al., 2009).

In some cases, knowledge of trajectory and proximity information is sufficient to infer the activity state. The addition of a-priori information about the targets and their work packages, the location of regions of interest plus their meaning, and the trajectories of each target enables the decoding of activities through Markov models based on work process diagrams (Gong & Caldas, 2009; Yang et al., 2011). Site analysis is possible for earthwork processes since the quantity and types of equipment are limited during this part of the construction phase, and the activities can be inferred from the interaction dynamics of the machines (Ogunmakin et al., 2013).

Productivity Estimation. Analysis of activity state estimates over time, when connected to specific work packages, provides productivity data for the work packages (Gong & Caldas, 2011). Over short time intervals, with specific work packages, productivity can be inferred through the activity states coupled with some minimal information regarding the task (Navon, 2005). For longer time intervals, however, it is more useful to connect activity statistics to actual progress, which requires progress tracking. To date, we are unaware of any vision-based solutions that provide productivity estimation of construction site operations through the measurement of progress and activity states over time.

3 VOLUME CALCULATION USING PHOTOGRAMMETRY

The proposed approach to quantify the amount of excavated soil on an excavation site is to create a 3D point cloud of the site space via photogrammetry. The photogrammetry algorithm used is VisualSFM (Wu, 2011). The algorithm locates feature points within the recorded photos using the scale invariant feature transform SIFT (Lowe, 1999). Those features are then matched among the individual photographs. Features visible in at least three photographs are then used to triangulate points in three dimensional space (see Figure 6a). Those points eventually form a point cloud. Additionally the patch based multi view stereo (PMVS) algorithm (Furukawa & Ponce, 2010) can be used to create a denser representation of the scene and to add information about the local surface normals throughout the point cloud.

Once the point cloud is generated, volume calculation of the point cloud required several additional steps. First, the points outside of the excavation area are cleared from the point cloud through conditional Euclidean cluster analysis (Lefkovitch, 1980). Second, a consistent top plane, which covers the excavation area, is found by performing a vertical histogram analysis, showing a sudden drop in the number of points along the vertical axis, which is consistent over the time series of point clouds in the case studies.

For sites with very close surrounding trees or buildings the histogram might not be sufficiently accurate, in which case markers are placed on site. Those markers act to precisely demarcate a reference plane, most commonly the ground level of the soil prior to excavation, for volume calculation. To resolve the scale ambiguity of the site arising from photogrammetry, the scale of the site is calculated from known distances between marker points, or known distances between points in the scene. Either the known features, such as bored pile walls, or the markers have been located within the point clouds in order to determine the scale. All points above the top cover plane are removed and gaps remaining in the point set are filled using linear interpolation. The volume of the excavation pit is calculated using signed tetrahedron volumes of a mesh generated from the top plane and the point cloud. More details on the entire procedure are given by Bügler et al. (2013). The documented point clouds in this paper were each calculated on the basis of several hundred photos, per session, processed offline to create a point cloud of high density, containing more than a million points.

4 VIDEO ANALYSIS

Following the approach by Ogunmakin et al. (2013), this paper presents an automatic surveillance system for processing videos of excavation operations that requires *a-priori* information about the construction work site layout and any activities of interest. The problem setup involves a single monocular camera configured to view a scene where activities of interest could occur. The camera view is assumed to provide an angled sideways view from a high enough vantage point that the region of interest fits within the field of view. The targets of interest are dump trucks and excavators. After processing the video, the system provides an automated report of the activity states of the targets (moving, static, or excavating) and the statistics associated with the activities of interest.

Figure 2 depicts the main steps of the online algorithmic processing, which includes the following elements:

1. Background modeling: Background estimation to initialize the background subtraction model and to identify potential targets via foreground detection.
2. Target detection: Foreground objects analysis to detect new targets and learn their respective appearance models for re-identification.
3. Tracking: Check whether a tracker has been lost, check the entrances for new targets, and, if a target is detected, re-identify previously seen targets.
4. Activity status estimation: Activity status estimation for each tracked target based on regions of interest.

The system processes video at real-time, given a frame rate of 10 frames per second. Once the whole video sequence is tracked, the results are passed to an event detection processor that outputs the average time spent in the region of interest, the number of targets that entered the region of interest, the number of targets that entered the scene, how long they spent in each region of interest, and each target's deviation from the average time spent in the region of interest. The post-processing event detection step completes in less than 10 minutes of time for

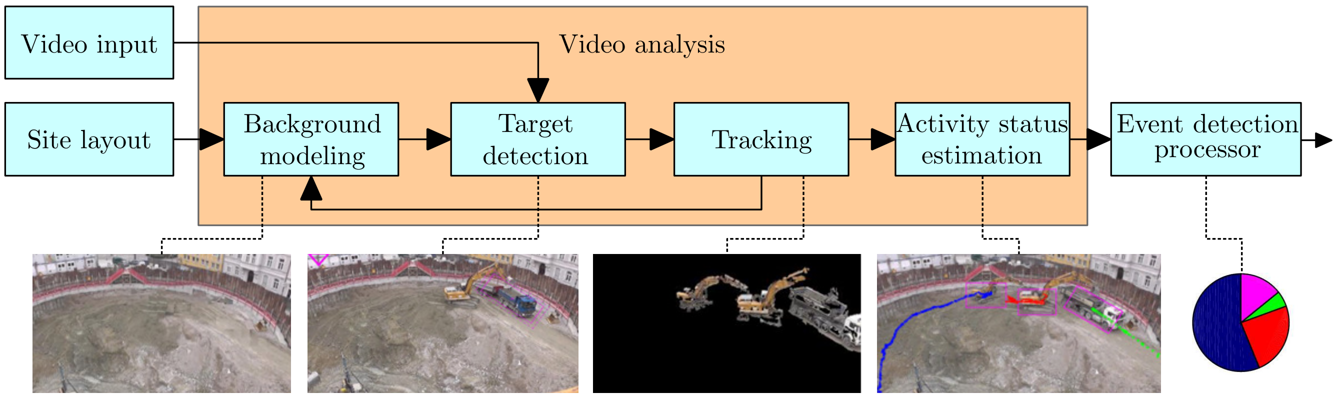


Figure 2: Process flow for the automatic surveillance system.

a full workday. The subsections that follow describe the different modules included in the system and their implementation details.

4.1 Foreground-Based Target Detection

Target detection is performed using the background Gaussian Mixture Model (GMM) method proposed by Stauffer & Grimson (1999), which determines foreground based on the probability of a pixel belonging to the background. The GMM is initialized using the background estimation technique of Reddy et al. (2009). Figure 3 illustrates different background models created for one of the case studies. The blurred or pixelized areas in the image were interpolated image regions, in order to remove machines from the scene.

Regions with low probabilities, as determined by a threshold, define the foreground regions. Coupling these foreground detections with the worksite layout information (e.g., entrance gates or zones), establishes when a target first enters the scene. For a given frame, the foreground regions need to be classified as excavator, dump truck, or neither. To differentiate between the two machines, the size, and aspect ratio of the detected entity is used to decide whether it is a dump truck or an excavator. The area of the region is used to determine whether it should be classified as one of these machines or as a spurious detection.

4.2 Kernel Covariance Tracking

An improvement on the kernel covariance tracker used in Ogunmakin et al. (2013) is utilized for tracking detected targets. Two improvements were made: (1) reduction of data before tracking (Kingravi et al., 2013), and (2) introduction of a scale space search with upper limits and lower limits. The data reduction step saves memory and lowers the computational cost of tracking. The scale space search allows the tracker to handle changes in scale. To initialize a tracker, the target's tracker model is learnt by mapping its feature vector, consisting of its color and spatial information, into a higher dimensional space using the Gaussian kernel and performing kernel principal component analysis. For every frame and each target, a gradient ascent procedure localizes the target by comparing the foreground image data with the targets' learnt model in order to optimize the region similarity.

4.3 Activity Status Estimation

The activity status of the machines follows that of Bügler et al. (2014), where machine activity is decomposed into *static*, *moving*, *absent*, and *filling*. An activity check is performed when an excavator and a dump truck are in close proximity. This is the case when the closest distance between the outlines of the machines is low relative to the sizes of the machines. In the presented case studies, a threshold equal to the expected length of one dump truck proved effective. Then, the movement of the excavator in the proximity zone of a dump truck establishes when an excavator might be filling a dump truck. This state can only be triggered when the two machines are in close proximity and the dump truck is static. Compared to Golparvar-Fard et al. (2013), where the machine activity is determined using machine learning with spatio-temporal features, the presented system utilizes optical flow to detect the dump-truck filling activities. It relies on the angled sideways camera view (as opposed to a top-down view).

4.4 Event Detection Processor

The event detection processor takes as input the trajectory information from the tracker and the results from the activity status estimation, and uses this to generate the statistics needed to determine the time spans of the work activities of the excavators and dump trucks. The metrics computed are the number of dump trucks that entered the scene (n_{trucks}), how much time they spent in the scene (t_{scene}), how much time they spent in the region of interest getting filled (t_{roi}), how many bucket loads of

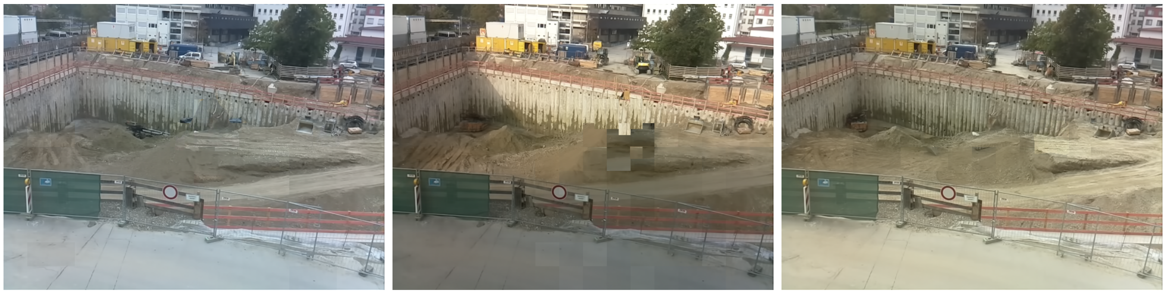


Figure 3: Estimated background models (pixellated regions are interpolation artifacts from deleted machines).

soil were placed in each dump truck ($n_{buckets}$), and how long the machines spent idle while in the scene (t_{idle}).

The number of dump trucks that entered the scene, n_{trucks}, is computed by counting the number of trackers initialized for detected dump trucks. Their duration in the scene, t_{scene}, is obtained by subtracting the time stamp they left the scene from the time stamp they entered the scene. The time spent being loaded by the excavator, t_{roi}, is determined using the results from the activity estimation. It sums up the total amount of time the activity estimation detected that the dump truck was being loaded. The number of bucket loads, $n_{buckets}$, is also determined by the activity estimation module by counting how many times it detected the excavator bucket over the dump truck using optical flow (optical flow measures the apparent motion of pixels from one frame to the next). The time the machines spent idle in the scene, t_{idle}, is determined by checking how much movement has happened between frames. Movement below a threshold triggers the idle state.

The event processor tabulates the temporal statistics of the activities and also identifies events, such as filling cycles and outlier time spans. A sample activity timeline and summary pie chart is given in Figure 4 for a 6 minute segment of processed video.

5 PERFORMANCE ANALYSIS

The performance of machines, operators, workers, and processes is commonly described by performance factors, such as soil removed per hour/day, the quantity of dump truck loads per hour, or the time required by a truck to transport soil to a dump site. The data acquired by cameras, as described in the previous sections, may indicate instances where those factors deviate from established values. In those cases it is important to determine the reasons for the failure and update the respective performance factors.

In the context of this research the performance is defined as the amount of soil removed from the site per time. Given two volume measurements provided by the photogrammetry approach, this is calculated as defined in Equation (1), where t_1 and t_2 are the points in time and are v_{t_1} and v_{t_2} are the respective volumes of the excavated pit,

$$\tau_p = \frac{\Delta v}{\Delta t} = \frac{v_{t_2} - v_{t_1}}{t_2 - t_1} \tag{1}$$

This performance factor is calculated on a daily basis, but it does not provide deeper estimates during the course of the day, nor insight into the causes for the observed performance factors. Thus, video analysis is used to calculate the performance factor on a hourly (or finer) basis based on the number of excavation buckets $n_{buckets}$ counted. The volume of a bucket is assumed known for a given excavator. The automatically tabulated statistics associated to daily activity levels and provided via charts assist with the interpretation of the measured productivity levels.

Due to the swell factor of soil, which expands during excavation, and due to the varying filling level of the bucket, activity analysis only provides a rough estimate and the resulting performance factor τ_v can be assumed to be larger than τ_p. Furthermore, the volume error will accumulate on the long run and reduce the reliability of the volume estimates. As the swell factor and the filling level of the bucket can be assumed to have a low variance within a day or so, the statistical data of the video analysis is normalized by the daily volume measurement, which provides an absolute measure of the excavated soil. Normalization is achieved by assuming that both measurements should yield an equal performance factor. Imposing equality involves scaling τ_v to the same value as τ_p by correcting the bucket volume to the mean volume required to equalize the performance factors. The corrected bucket volume is calculated by dividing the excavated volume calculated via photogrammetry by the number of bucket loads counted via video analysis. Taking this corrected bucket volume into account, the statistics resulting from the video analysis can be accurately scaled to yield hourly performance factors for the excavation process that do not suffer from a cumulative error.

By combining the absolute volumetric measurements of photogrammetry with the fine temporal resolution of video analysis, the presented approach automatically pro-

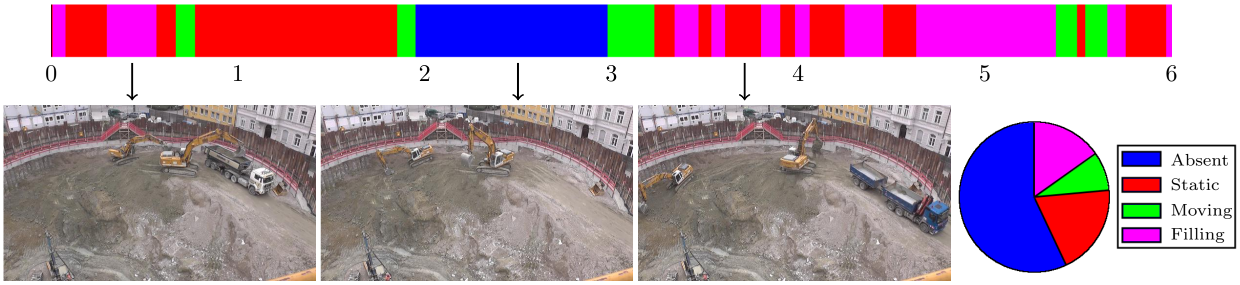


Figure 4: Dump truck state estimates for a video segment of 6 minutes duration, plus activity states in a pie chart.

vides accurate earth excavation performance factors for arbitrary time intervals. In addition, the developed video analysis techniques allow for a detailed analysis of the causes for performance fluctuations. As the event detector allows for measuring time intervals, a detailed analysis of the interaction between excavator and dump trucks becomes possible: If the created statistic shows that the dump trucks are absent over long periods of time, the reason for a low productivity is an insufficient number of available dump trucks. If on the other hand, the absent time is reasonably low and the filling time is higher than expected, the low productivity is caused by the excavator which can result from tough earth conditions, for example.

On the basis of these statistics and their root causes, the site manager can take the appropriate measures, i.e. employ more dump trucks or use alternative excavation machines. The resulting data can also be used for updating the performance factors database most construction companies maintain as a basis for project calculation and tendering.

6 CASE STUDIES

This section outlines two large construction projects that have been used to perform case studies using the proposed approach. Both projects included very large excavation tasks which were tracked using video cameras and photogrammetry. For the video analysis, an accompanying annotation software package was created for manual generation of the ground truth. An operator manually counted buckets and dump trucks using this software package, as well as annotated the activity states. The operator was shown a time-lapse of the video while clicking a button for bucket loads, as well as for entering or leaving dump trucks. All averages provided are robust averages, which limit the influence of outliers on the mean calculation, based on robust regression (Huber & Ronchetti, 2009). Points farther from the predicted linear regression models are given lower weights in each iteration until the model converges. The weights are thresholded to determine which data points are outliers.

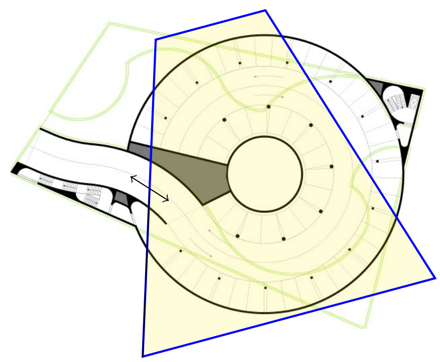


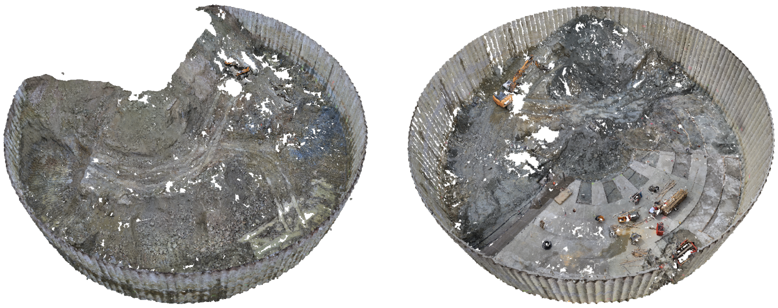
Figure 5: Plan view of the Josephsplatz construction site with camera view overlay (opaque trapezoid).

6.1 Underground Parking Garage Josephsplatz

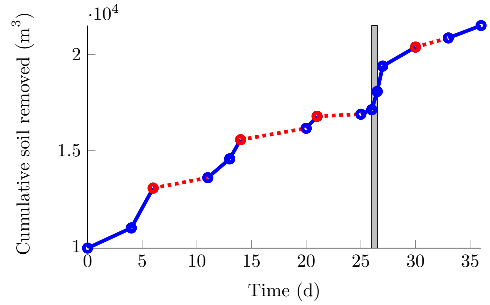
The first case study is a parking garage being built in plaza Josephsplatz within Munich, Germany. The layout of the site is illustrated in Figure 5, with an overlay of the region sensed by the video camera (trapezoidal region) and a double-arrow indicating the entrance/exit region of interest for the dump-trucks. Objects leaving the field of view through other areas were labeled lost but not gone. Frames of the recorded video are shown in Figure 4.

Photogrammetry. Photos to track the progress were taken throughout the excavation process, which was observed for 36 days. Point clouds were generated at intervals of 2-6 days by an operator with a single mobile camera. The limiting factors preventing more frequent point cloud estimates were holidays, available manpower, and weather conditions. The resulting point clouds were then used to measure the volume between measurement time points. The course of the excavation process is illustrated with two point clouds in Figure 6a. Generally point clouds could be created on a daily basis to give feedback about the previous day's productivity on each morning. This will give decision makers the opportunity to react appropriately to observed delays.

As video recording of site operations occurred on day 26 for a continuous 4 hour time period (the gray region in Figure 6b), point cloud measurements were specifically made to coincide with the video recording time-span. Two

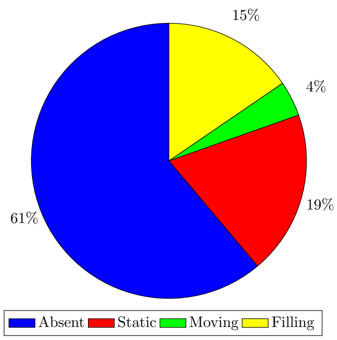


(a) Point clouds with color.

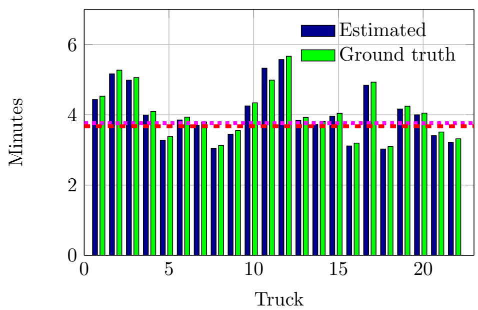


(b) Cumulative excavated volume.

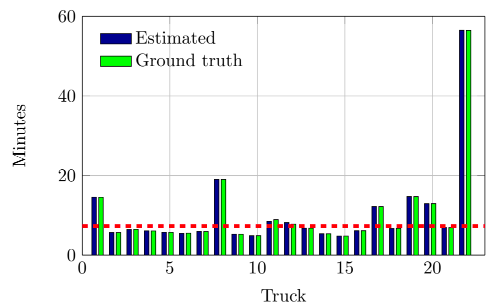
Figure 6: Photogrammetric progress monitoring results at Josephsplatz site during the observation period.



(a) Dump truck states.



(b) Loading time per truck.



(c) Inter-arrival times between dump trucks.

Figure 7: Dump truck activity statistics for Josephsplatz site on day 1.

additional point clouds were generated prior to and after the video was recorded. For the recorded four hour period an excavated volume of $\Delta v = 418m^3$ was calculated, yielding an average performance factor of $\tau_p = 104.5\frac{m^3}{h}$.

Video Analysis. After the video tracking process completed, the event processor combined the tracking and activity estimation results to generate the illustrated statistics. A total of 22 trucks were detected in the processed video, with 0% error compared to the ground truth. A total of 171 bucket loads were detected compared to the ground truth of 177 bucket loads (3.4% error). The total volume of the excavator's bucket was $2.5m^3$. Using the ground truth, this results in a performance factor of $110.63\frac{m^3}{h}$. Using the estimated 171 buckets, this results in a performance factor of $106.88\frac{m^3}{h}$. The constants entering in these calculations were re-estimated based on the photogrammetry estimates. Calculating the corrected volume of soil excavated per bucket, using the ground truth, results in a bucket volume of $2.36m^3$. Calculating the corrected volume of soil excavated per bucket, using the estimation, results in a volume of $2.44m^3$.

Figure 7b shows the total time each dump truck spent in the scene for loading, compared to the manually observed ground truth. The red line indicates the estimated average loading time per dump truck, excluding outliers, which was 3.67 minutes, while the pink line indicates the ground truth average time per dump truck which was 3.76 minutes, for a 2.4% error between the estimated results and ground truth. There were five identified outliers taking 4.8 minutes or more. The estimated total amount of time the dump trucks were in the scene was 88.3 minutes out of 240 minutes of video (39% of video duration), compared to the ground truth of 89.9 minutes (37.5% of video duration), with 1.7% error between the estimated results and ground truth. The inter-arrival time is the time duration between when the previous dump truck enters and the next dump truck enters. The inter-arrival time for this video sequence is shown in Figure 7c. For the first dump truck, the inter-arrival time quantity measures the amount of time from the start of the video to when the dump truck first entered. The estimated average inter-arrival time between the dump trucks was 7.3 minutes excluding the automatically identified outliers, while the ground truth was deviating by 0.02%.

Productivity Assessment. Using the bucket capacity, corrected using the photogrammetry measurements, of $2.36m^3$ together with the detected bucket load provides soil removal estimates at finer time intervals. Figure 8 shows a plot of the cumulative soil removed over the 240 minutes of recorded video compared to the ground truth (right y-axis labels). Measurements are provided for each ten minute interval. Estimates of the performance factor are provided for each thirty minute interval (left y-axis labels). The chart provides information on the overall productivity of the excavator and dump-truck collaboration. Productivity was steady to the 3 hour mark, then stalled. Additional statistics to provide context for the productivity levels are provided below.

The excavation volume measured over time by means of photogrammetry is plotted in Figure 6b. The progress was thresholded to determine when significant progress occurs. Blue solid line segments indicate a time interval where significant progress occurred, while red dashed line segments indicate a time interval where little progress occurred. As seen in Figure 6b, there were four measurement intervals where there were no significant progress. The average amount of soil removed per day during these intervals was less than the threshold of $200m^3$ (a different level could be chosen). The average volume removed per day, excluding days without significant progress is $874m^3$. Presuming at least an 8 hour work day, the productivity seen on day 26 appears to be about average.

The statistical evaluation of the video analysis results is depicted in Figure 7a. It indicates what percentage of total recording time the different activity states of the dumper trucks were observed. The aggregate statistics indicate that it is possible to improve on the efficiency of the process by incorporating more dump trucks to reduce the idle times of the excavator and increase the amount of soil removed from the site, since only 39% of the available filling time was used. Analyzing further, the fairly steady arrival of dump trucks, with the exception of the few outliers, indicates that there were no transportation issues. The average loading time, 3.76 minutes, and the average inter-arrival time, 7.3 minutes, also show that it's possible to incorporate additional dump trucks. The inter-arrival times did increase in the latter half of the day, the source of which should be investigated further by the project manager. The productivity levels seen during this period could be improved.

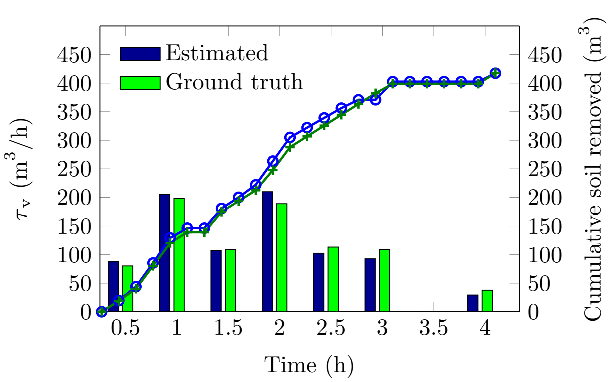


Figure 8: Soil removal statistics for Josephsplatz site

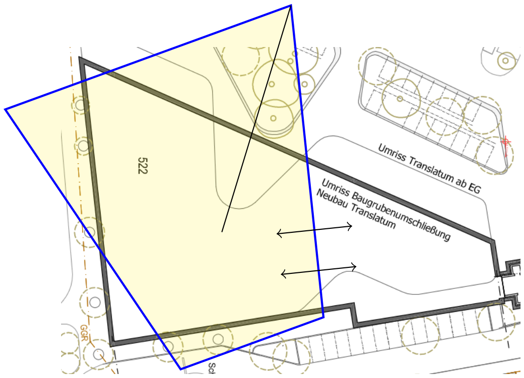
6.2 Hospital Building translaTUM

The second project is a building called *translaTUM* which is an extension of the Hospital "Klinikum rechts der Isar" in Munich, Germany. The layout of the site is illustrated in Figure 9a, with an overlay of the region depicting the video camera perspective (trapezoidal region) and double-arrows indicating the entrance/exit region of interest for the dump-trucks. Site observation involved six pan-tilt cameras mounted around the site and used to take photographs. One camera was configured to record and transfer video. Additional photos were taken by an operator with a single mobile camera. Video was recorded for four full days, from day 8 to day 11, while the point clouds were created every evening. Days 9 and 10 fell on the weekend, during which no work was performed. A snapshot of the recorded video is shown in Figure 9b.

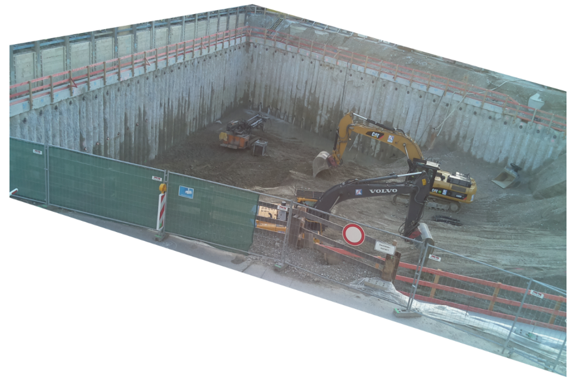
Photogrammetry. Due to rain, snow and difficult light conditions, not every photo session had sufficient quality to generate a clean point cloud. Therefore, only the point clouds with sufficient density were analyzed Eleven measurement sessions out of 90 were of satisfactory quality. The low success rate can be fully attributed to adverse weather conditions which produced low quality images during periods of strong rain and to improper placement of some cameras relative to the active work zone. Strong winds shaking the cameras additionally blurred the photos taken. When used by construction industry, these shortcomings would be counteracted by investing in better hardware. Two of the generated point clouds are shown in Figure 10a, and a time series plot of the cumulative excavated volume is graphed in Figure 10b. The larger time gaps between point cloud measurements have flat progress indicating that construction activity was also impacted by the same weather conditions.

During the video recording period, marked by the gray region in Figure 10b, a total of $\Delta v = 774m^3$ were excavated as calculated through the photogrammetry approach. This results in a performance factor of $32.25\frac{m^3}{h}$.

Video Analysis. Video recorded on work days 8 and 11 was processed. The ground truth was again manually annotated for the activity information sought. On day 8, 37 out of 38 trucks were detected by the system. The system detected 305 out of 313 bucket loads (2.56% error). On day 11, the system detected 74 trucks when there were actually 73 trucks, and 334 out of 596 bucket loads (55.04% error). The total volume of the excavator's bucket was $1m^3$. Using the ground truth of 909 buckets, results in a performance factor of $37.88\frac{m^3}{h}$. Using the estimated 639, results in a performance factor of $26.63\frac{m^3}{h}$. The measurement errors on day 2 lead to a 29.70% error in the performance factor. Calculating the corrected volume of soil excavated per bucket, using the ground truth, results in a bucket volume of $0.85m^3$. Calculating the corrected volume of soil excavated per bucket, using the estimation, results in $1.21m^3$. The performance factor er-

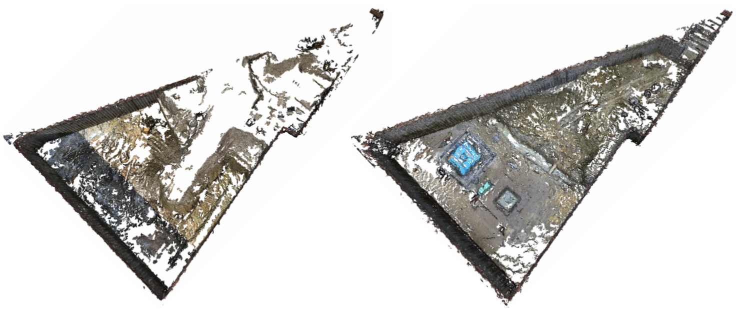


(a) Plan view with camera view overlay (opaque trapezoid).

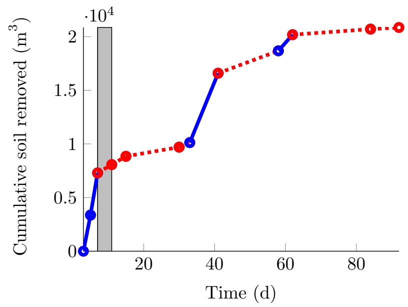


(b) Cropped snapshot from recorded video.

Figure 9: The translaTUM construction site information, video geometry, and image view.



(a) Point clouds with color.



(b) Cumulative excavated volume.

Figure 10: Photogrammetric progress monitoring results at translaTUM site during the observation period.

ror leads to a larger, unrealistic excavation volume for the bucket, so the value was capped at $1m^3$.

Figure 11 contains charts of the amount of time each truck spent in the scene as part of the loading process, for the two days analyzed. The average estimated loading time for day 8 is 4.87 minutes, with a ground truth time of 4.23 minutes, for a 15.1% error. The average estimated loading time for day 11 is 3.03 minutes, with a ground truth loading time of 3.85 minutes, for a 21.30% error.

The major outliers over the average red line in Figure 11a were due to the tracker losing the targets and drifting to another foreground object in the scene. Track loss was caused by the excavator moving behind a fence, as shown in Figure 9b, whereby the truck too would drive to a region of space occluded by the fence. Such errors can be prevented by better camera placement, or through multiple cameras. The major outliers below the average red line in Figure 11c were due to occlusion-based track loss. The detector detected trucks it had missed when they first entered the scene, and it also re-detected the trucks the tracker lost as they exited the scene.

Figure 12 charts the inter-arrival times of the trucks tracked as well as the ground truth (the red line is the average value). The average estimated inter-arrival time for day 8 is 6.07 minutes, with a ground truth time of 6.24 minutes, for a 2.72% error. The average estimated inter-arrival time for day 11 is 6.07, with a ground truth time of 6.4 minutes, for a 5.44% error.

Productivity Assessment. Using the bucket capacity of $1m^3$ obtained from capping the corrected bucket capacity v_{bucket_cor} to the maximum possible, together with the detected bucket loads over time provides the soil removal statistics depicted in Figures 14 and 15. Figure 14 shows the estimated amount of soil removed per hour on both days, and Figure 15 shows the cumulative soil removed on both days. For the first day the algorithm does a good job matching the ground truth. Due to near total occlusions of one of the excavators on the second day, many of the shovel loads were missed. The estimated volume excavated is 639 m^3 which is 17.44% off from the photogrammetry measurement. In both cases, consideration of the slopes without regard to the actual values provides a high-level indication of productivity throughout the day. On Day 8, productivity was quite good up until the fourth hour (240 minutes in) at which point it practically halted. Day 11 productivity experienced a lull after about 5 hours in (300 minutes), then resumed for about 3 hours.

The progress of the excavated volume over time is plotted in Figure 10b. Blue solid line segments indicate time intervals with significant progress, while red dashed line segments indicate time intervals with low progress. As seen in Figure 10b, there are six measurement intervals with no significant progress (less than $200m^3$). This is relevant because excavation processes, especially of deeper excavation pits, require shoring operations, such as the anchoring of bored pile walls. These operations then temporarily interrupt the excavation processes. The average volume removed per day, excluding days without significant progress, is $913m^3$. As a performance factor, this equates to $91.3m^3/h$ for a 10-hour day. By day 31, around 10,000 m^3 of soil was removed, which tracks closely to the soil removal from the Josephsplatz sites, however, there were more days with little to no activity compared to Josephsplatz. For the two days observed, the productivity levels were below the productivity level associated to the typical day. Even if day 8 were to have been fully utilized, progress would be below average for that day.

The event processor generated the pie charts in Figure 13 of the aggregate statistics for the two days. The statistics for the two days are similar when considering the amount of time that dump trucks were absent versus within the scene. Much like the Josephsplatz case, the high percentage of dump truck absence means that additional dump trucks should improve progress. Further, the high percentage of dump-truck absences and static moments combined with the disparity between the dump-truck inter-arrival times and the dump-truck filling times indicates that performance could be nearly doubled to match the average performance value.

6.3 Discussion

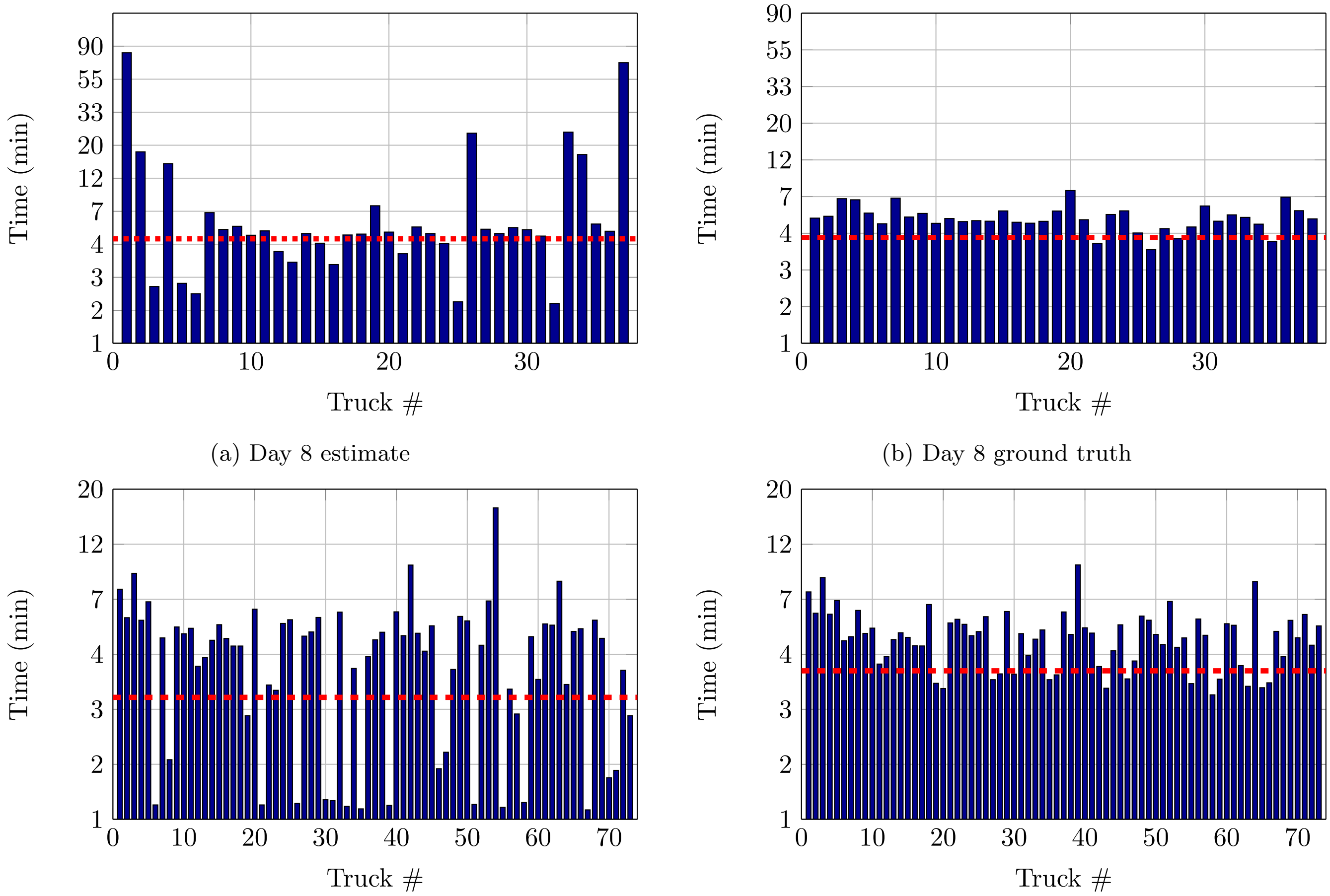
Combining photogrammetry and automated video analysis technologies overcomes their individual limitations and enables earthwork productivity analysis with sufficient accuracy and temporal resolution so as to provide daily feedback to project managers. While photogrammetry provides precise information about the quantity of removed earth volume, its application demands comparatively high effort and is thus performed only once a day. Productivity analysis based purely on photogrammetry will thus have a rather coarse temporal resolution. In addition, it is not possible to identify the probable causes for low performance. On the other hand, video analysis provides in-depth insight into the interaction between excavators and dump trucks, and allows statistical analysis of processing times, idle times and more, at a finer temporal sampling rate relative to photogrammetry. However, video analysis does not provide precise quantities of excavated earthwork due to measurement uncertainty. Coupling the two measurement strategies results in a powerful tool which allows earthwork productivity analysis on a fine temporal resolution. The daily progress measurements from photogrammetry are converted to productivity estimates at finer time intervals within the day, and also made available on a daily basis.

Productivity estimates, whether high or low, do not necessarily provide the full story as to what variables may be leading to the estimated values. The output statistics associated to the automated video analysis provide ancillary information to help identify the sources of low productivity, as well as to confirm the reasons behind high productivity. The statistics and charts can enable the site manager to identify the reasons underlying the productivity fluctuations. In particular, by providing support information for more detailed process analysis, the system can help identify the causes of potentially low performance and contemplate interventions to improve productivity.

For example, long excavator idle times and low bucket loads indicate low dump truck availability. Conversely, high dump truck idle times and relatively low bucket loads imply lower filling rates relative to dump truck arrival rates, pointing to insufficient excavator availability or low removable soil volume. As a final example, low dump truck idle times, high quantities of bucket loads, and high productivity imply a smoothly running operation. Usually dump trucks are hired on a daily basis and excavators also take time to be transported to and from the site. Therefore providing feedback of the prior day's productivity and activity levels to decision makers each morning is sufficient to allow countermeasures to be put into place.

For both methods, however, there are circumstances that prevent the generation of accurate measurements. In the case of photogrammetry, the main problems arise when the excavation area is covered by snow or when large pools of water form due to heavy rain. The missing measurements do not necessarily result in the inability to measure progress, as subsequent measurements on days with favorable weather still capture progress to date and can be fused with the automated video analysis. The work presented two case studies that showed that photogrammetry is able to properly document the time periods with high productivity, thus any missing measurements derived from visually unfavorable conditions do not cause problems.

Due to the finer time scale of the operations to be monitored for the visual activity surveillance thread as well as the instantaneous nature of the information to measure, the activity algorithm output is less robust to poor visual conditions, such as occlusions. However,an activity monitoring algorithm does not need to be 100% reliable to provide useful information. The use of robust statistics leads to estimates that are close to the ground truth in spite of the uncertain measurements. During good visual conditions (as seen in the first case study) the productivity estimation error was 10% or less. In the second case study, poor visual conditions created by the occlusion of certain periods of excavation by a fence, led to higher errors. For excavation processes that maintain levels of operations consistent with those found in this paper, the extracted statistics and charts are sufficient to identify outliers and to concentrate attention on specific times of the day. The main conclusion is that slightly erroneous continuous monitoring of excavation operations should be



(a) Day 8 estimate

(b) Day 8 ground truth

(c) Day 11 estimate

(d) Day 11 ground truth

Figure 11: Time spent per dump truck in the scene while being filled.(Log scale)

preferred to coarse work sampling methods that may not properly capture on-site activities.

Although the vision-based algorithms may indeed miss certain activities due to occlusions, there is still valuable information that can be gleaned from the available data. In the one scenario with poor visual conditions, the continuous monitoring was able to identify idle and high productivity periods during the day. Although the rates were incorrect and led to a poor estimate of the true productivity on Day 11 of the TranlsaTUM site (see Figure 14), the rates still correlated with the actual rates. Work sampling may not capture these time-variable statistics as finely as continuous monitoring can. Furthermore, the other quantities such as the number of trucks detected during the day, and the loading time, serve as a sanity check on potentially erroneous rates. However, in industrial use cases, the shortcomings can easily be counteracted by investing in better hardware and having a more informed setup. Making sure to capture the scene from angles where the interaction of the involved machines is clearly visible without occlusions can strongly reduce the mentioned problems.

The presented concept can help decision makers on site to react to upcoming problems in different ways. When dump trucks are stuck in traffic, the manager could decide to reduce the number of excavators on the site to match the rate of arriving dump trucks, or to increase the number of trucks. When dump trucks queue on site due to low excavator productivity, more excavators could be deployed, or if the site cannot fit more excavators, the productivity increase by additional excavators is not expected to be worthwhile, or if there are no additional excavators available, the number of dump trucks could be reduced to match the excavator performance.

7 CONCLUSION

This paper integrates two vision-based sensing methods for generating productivity information regarding excavation operations. Both static and mobile cameras were used to sense the site over extended periods time. The first method involved photogrammetry to generate volume measurements, while the second involved the tracking of excavation processes involving excavation and hauling equipment. Together, they serve to provide fine timescale estimates of productivity (e.g, hourly or finer). Fur-

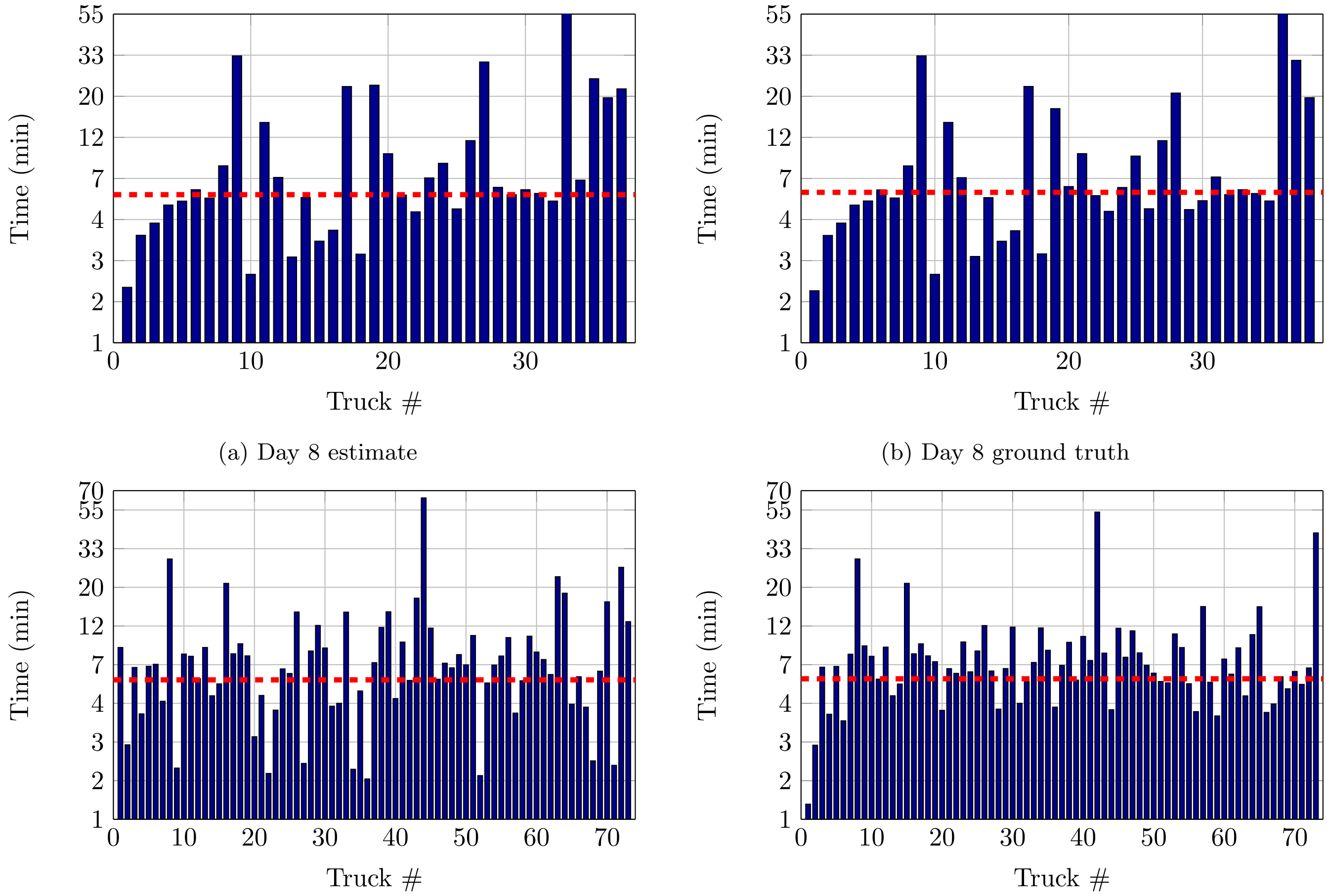


Figure 12: Inter-arrival times between trucks. (Log scale)

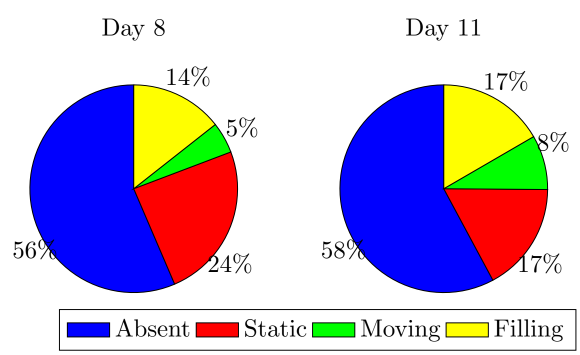


Figure 13: Aggregate statistics for translaTUM.

thermore, the daily progress levels, activity states, and automatically extracted statistics assist with the interpretation of the productivity levels estimated. Future research should address the extension of the approach towards the combination of data recorded by multiple video cameras on-site to decrease the likelihood of occlusions.

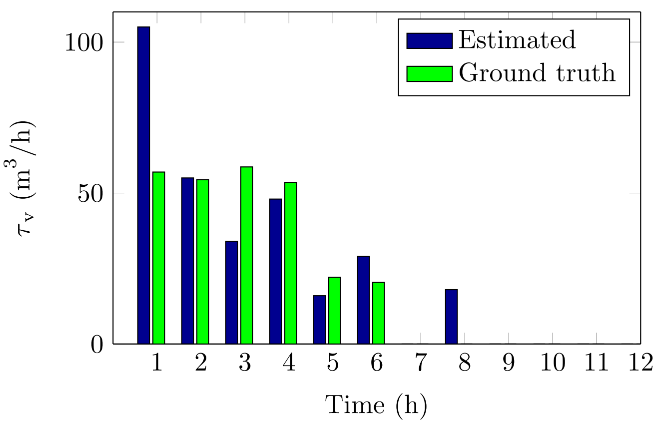
ACKNOWLEDGEMENTS

The research presented in this paper was supported by grants of the US National Science Foundation (NSF) under grant agreement #1030472 and the Bavarian Research Foundation (Bayerische Forschungstiftung) under the research project FAUST. The authors would also like to thank the industrial partners for their support, in particular those who provided access to construction sites for data collection. The authors would furthermore like to thank Daniel H. Biedermann for the fruitful discussions. All results and conclusions presented are that of the authors alone.

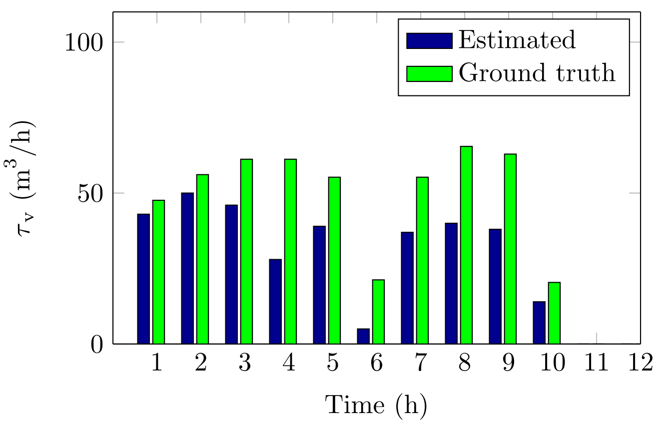
REFERENCES

Ake, D. (2001), Laser receiver and angle sensor mounted on an excavator, Patent US 6,263,595.

Assaf, S. A. & Al-Hejji, S. (2006), Causes of delay in large construction projects, *International Journal of Project Management*, 24(4), 349–357, doi:10.1016/j.ijproman.2005.11.010.

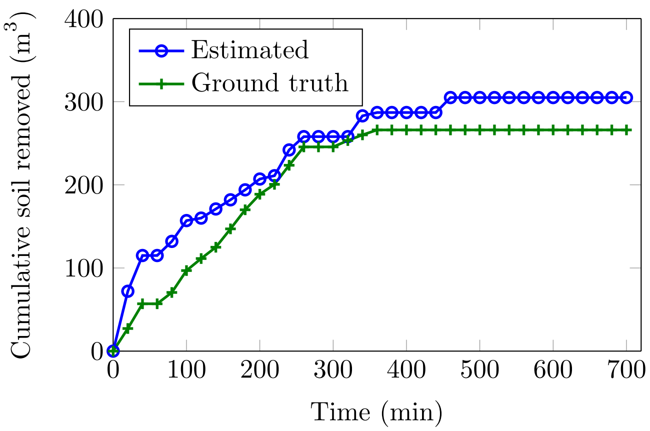


(a) Day 8 estimate and ground truth

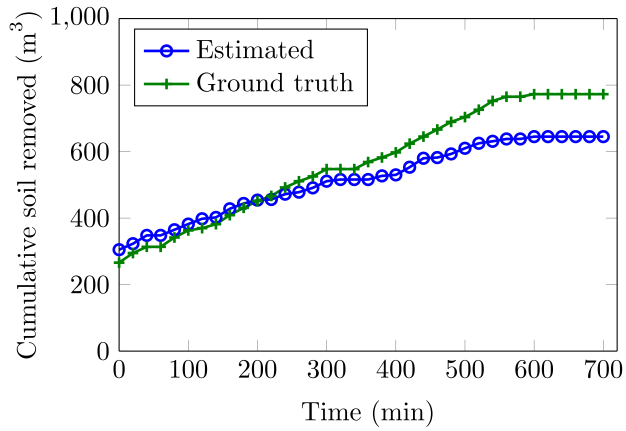


(b) Day 11 estimate and ground truth

Figure 14: Performance factor over time.



(a) Day 8 estimate and ground truth.



(b) Day 11 estimate and ground truth.

Figure 15: Estimated cumulative soil removed per 20 minute interval.

Aydin, C. C. (2014), Designing building façades for the urban rebuilt environment with integration of digital close-range photogrammetry and geographical information systems, *Automation in Construction*, 43, 38–48, doi: 10.1016/j.autcon.2014.03.005.

Bohn, J. & Teizer, J. (2010), Benefits and barriers of construction project monitoring using high-resolution automated cameras, *ASCE Journal of Construction Engineering and Management*, 136(6), 632–640, doi: 10.1061/(ASCE)CO.1943-7862.0000164.

Bosché, F. (2010), Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction, *Advanced Engineering Informatics*, 24(1), 107–118, doi: 10.1016/j.aei.2009.08.006.

Braun, A.; Tuttas, S.; Borrmann, A. & Stilla, U. (2015), A concept for automated construction progress monitoring using BIM-based geometric constraints and photogrammetric point clouds, *Journal of Information Technology in Construction*, 20(8), 68–79.

Brilakis, I.; Park, M.-W. & Jog, G. (2011), Automated vision tracking of project related entities, *Advanced Engineering Informatics*, 25(4), 713–724, doi:10.1016/j.aei.2011.01.003.

Bügler, M.; Metzmacher, H.; Borrmann, A. & van Treeck, C. (2013), Integrating Feedback from Image Based 3D Reconstruction into Models of Excavation Processes for efficient Scheduling, in *Proceedings of the EG-ICE Workshop on Intelligent Computing in Engineering*, doi: 10.13140/RG.2.1.4105.9684.

Bügler, M.; Ogunmakin, G.; Teizer, J.; Vela, P. A. & Borrmann, A. (2014), A Comprehensive Methodology for Vision-Based Progress and Activity Estimation of Excavation Processes for Productivity Assessment, in *Proceedings of the EG-ICE Workshop on Intelligent Computing in Engineering*, 1–10, doi:10.13140/RG.2.1.4630.2561.

Cheng, T.; Teizer, J.; Migliaccio, G. C. & Gatti, U. C. (2013), Automated task-level activity analysis through fusion of real time location sensors and worker's thoracic posture data, *Automation in Construction*, 29, 24–39, doi: 10.1016/j.autcon.2012.08.003.

Chi, S. & Caldas, C. H. (2011), Automated Object Identification Using Optical Video Cameras on Construction Sites, *Computer-Aided Civil and Infrastructure Engineering*, 26(5), 368–380, doi:10.1111/j.1467-8667.2010.00690.x.

Collins, R. T.; Lipton, A. J.; Kanade, T.; Fujiyoshi, H.; Duggins, D.; Tsin, Y.; Tolliver, D.; Enomoto, N.; Hasegawa, O.; Burt, P. & Wixson, L. (2000), *A System for Video Surveillance and Monitoring*, volume 823, doi: 10.1109/TPAMI.2000.868676.

Costin, A.; Pradhananga, N. & Teizer, J. (2012), Leveraging passive RFID technology for construction resource field mobility and status monitoring in a high-rise renovation project, *Automation in Construction*, 24, 1–15, doi: 10.1016/j.autcon.2012.02.015.

Dimitrov, A. & Golparvar-Fard, M. (2014), Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections, *Advanced Engineering Informatics*, 28(1), 37–49, doi:10.1016/j.aei.2013.11.002.

Fathi, H. & Brilakis, I. (2013), A videogrammetric as-built data collection method for digital fabrication of sheet metal roof panels, *Advanced Engineering Informatics*, 27(4), 466–476, doi:10.1016/j.aei.2013.04.006.

Furukawa, Y. & Ponce, J. (2010), Accurate, dense, and robust multiview stereopsis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), 1362–76, doi: 10.1109/TPAMI.2009.161.

Golparvar-Fard, M.; Bohn, J.; Teizer, J.; Savarese, S. & Peña Mora, F. (2011), Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques, *Automation in Construction*, 20(8), 1143–1155, doi:10.1016/j.autcon.2011.04.016.

Golparvar-Fard, M.; Heydarian, A. & Niebles, J. C. (2013), Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers, *Advanced Engineering Informatics*, doi: 10.1016/j.aei.2013.09.001.

Gong, J. & Caldas, C. (2009), *An Intelligent Video Computing Method for Automated Productivity Analysis of Cyclic Construction Operations*, 64–73, doi:10.1061/41052(346)7.

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(8), 1211–1226, doi: 10.1016/j.autcon.2011.05.005.

Goodrum, P. M.; Zhai, D. & Yasin, M. F. (2009), Relationship between Changes in Material Technology and Construction Productivity, *ASCE Journal of Construction Engineering and Management*, 135(4), 278–287, doi: 10.1061/(ASCE)0733-9364(2009)135:4(278).

Gouett, M. & Haas, C. (2011), Activity analysis for direct-work rate improvement in construction, *Journal of Construction Engineering Management*, 137(12), 1117–1124, doi:10.1061/(ASCE)CO.1943-7862.0000375.

Grau, D.; Caldas, C. H.; Haas, C. T.; Goodrum, P. M. & Gong, J. (2009), Assessing the impact of materials tracking technologies on construction craft productivity, *Automation in Construction*, 18(7), 903–911, doi: 10.1016/j.autcon.2009.04.001.

Haritaoglu, I.; Harwood, D. & Davis, L. S. (2000), W4: Real-time surveillance of people and their activities, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 809–830, doi:10.1109/34.868683.

Huber, P. & Ronchetti, E. (2009), *Robust Statistics*, Wiley, 2nd edition.

Javed, O.; Rasheed, Z.; Alatas, O. & Shah, M. (2003), KNIGHT: a real time surveillance system for multiple and non-overlapping cameras, in *Proceedings of the International Conference on Multimedia and Expo. ICME '03.*, doi: 10.1109/ICME.2003.1221001.

Karsch, K.; Golparvar-Fard, M. & Forsyth, D. (2014), ConstructAide : Analyzing and Visualizing Construction Sites through Photographs and Building Models, *ACM Transactions on Graphics (TOG)*, 33(6), 176–187.

Khosrowpour, A.; Fedorov, I.; Holynski, A.; Niebles, J. C. & Golparvar-Fard, M. (2014), Automated Worker Activity Analysis in Indoor Environments for Direct-Work Rate Improvement from Long Sequences of RGB-D Images, in *Construction Research Congress 2014*, 729–738, doi: 10.1061/9780784413517.075.

Kingravi, H. A.; Vela, P. A. & Gray, A. (2013), Reduced Set KPCA for Improving the Training and Execution Speed of Kernel Machines, in *Proceedings of the SIAM International Conference on Data Mining*, 441–449.

Koehn, E.; Young, R.; Kuchar, J. & Seling, F. (1978), Cost of delays in construction, *Journal of the Construction Division*, 104(3), 323–331.

Lefkovitch, L. (1980), Conditional clustering, *Biometrics*, 36(1), 43–58, doi:10.2307/2530494.

Lowe, D. (1999), Object recognition from local scale-invariant features, in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, IEEE, 1150–1157, doi: 10.1109/ICCV.1999.790410.

Lukins, T.; Ibrahim, Y.; Kaka, A. & Trucco, E. (2007), Now you see it: the case for measuring progress with computer vision, in *Proceedings of the 4th International SCRI Research Symposium*, 409–422.

Malinovskiy, Y.; Zheng, J. & Wang, Y. (2009), Model-free video detection and tracking of pedestrians and bicyclists, *Computer-Aided Civil and Infrastructure Engineering*, 24, 157–168, doi:10.1111/j.1467-8667.2008.00578.x.

Marzouk, M. M. & Zaher, M. M. (2015), Tracking Construction Projects Progress using Mobile Hand-Held Devices, in *Proceedings of the 5th International/11th Construction Specialty Conference*, doi:10.13140/RG.2.1.1968.7529.

Memarzadeh, M.; Heydarian, A.; Golparvar-Fard, M. & Niebles, J. C. (2012), Real-Time and Automated Recognition and 2D Tracking of Construction Workers and Equipment from Site Video Streams, in *Computing in Civil Engineering (2012)*, American Society of Civil Engineers, Reston, VA, 429–436, doi:10.1061/9780784412343.0054.

Mills, S. J.; Ford, J. J. & Mejías, L. (2010), Vision Based Control for Fixed Wing UAVs Inspecting Locally Linear Infrastructure Using Skid-to-Turn Maneuvers, *Journal of Intelligent & Robotic Systems*, 61(1-4), 29–42, doi:10.1007/s10846-010-9480-1.

Navon, R. (2005), Automated project performance control of construction projects, *Automation in Construction*, 14(4), 467–476, doi:10.1016/j.autcon.2004.09.006.

Ogunmakin, G.; Teizer, J. & Vela, P. (2013), Quantifying Interactions Amongst Construction Site Machines, in *Proceedings of the EG-ICE Workshop on Intelligent Computing in Engineering, Vienna, Austria.*

Park, M.; Jog, G. & Brilakis, I. (2011a), Initializing vision based trackers using semantic texton forests, in *Proceedings of the International Symposium on Automation and Robotics in Construction*, 379–384.

Park, M.-W. & Brilakis, I. (2012), *Enhancement of Construction Equipment Detection in Video Frames by Combining with Tracking*, 421–428, doi:10.1061/9780784412343.0053.

Park, M.-W.; Makhmalbaf, A. & Brilakis, I. (2011b), Comparative study of vision tracking methods for tracking of construction site resources, *Automation in Construction*, 20(7), 905–915, doi:10.1016/j.autcon.2011.03.007.

Peddi, A.; Huan, L.; Bai, Y. & Kim, S. (2009), Development of human pose analyzing algorithms for the determination of construction productivity in real-time, in *Proceedings of the Construction Research Congress*, 11–20.

Pradhananga, N. & Teizer, J. (2013), Automatic spatio-temporal analysis of construction site equipment operations using GPS data, *Automation in Construction*, 29, 107–122, doi:10.1016/j.autcon.2012.09.004.

Reddy, V.; Sanderson, C. & Lovell, B. C. (2009), An efficient and robust sequential algorithm for background estimation in video surveillance, in *Proceedings of the 16th IEEE International Conference on Image Processing (ICIP)*, 1109–1112.

Rezazadeh Azar, E. & McCabe, B. (2012a), Automated Visual Recognition of Dump Trucks in Construction Videos, *Journal of Computing in Civil Engineering*, 26(6), 769–781, doi:10.1061/(ASCE)CP.1943-5487.0000179.

Rezazadeh Azar, E. & McCabe, B. (2012b), Part based model and spatial–temporal reasoning to recognize hydraulic excavators in construction images and videos, *Automation in Construction*, 24, 194–202, doi:10.1016/j.autcon.2012.03.003.

Saidi, K. S.; Teizer, J.; Franaszek, M. & Lytle, A. M. (2011), Static and dynamic performance evaluation of a commercially-available ultra wideband tracking system, *Automation in Construction*, 20(5), 519–530, doi:10.1016/j.autcon.2010.11.018.

Siebert, S. & Teizer, J. (2014), Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system, *Automation in Construction*, 41, 1–14, doi:10.1016/j.autcon.2014.01.004.

Stauffer, C. & Grimson, W. (1999), Adaptive background mixture models for real-time tracking, *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, 246–252, doi:10.1109/CVPR.1999.784637.

Teizer, J.; Kim, C.; Haas, C. T.; Liapi, K. A. & Caldas, C. H. (2005), Framework for real-time three-dimensional modeling of infrastructure, *Transportation Research Record: Journal of the Transportation Research Board*, 1913, 177–186.

Teizer, J. & Vela, P. A. (2009), Personnel tracking on construction sites using video cameras, *Advanced Engineering Informatics*, 23(4), 452–462, doi:10.1016/j.aei.2009.06.011.

Tian, Y. L.; Brown, L.; Hampapur, A.; Lu, M.; Senior, A. & Shu, C. F. (2008), IBM smart surveillance system (S3): Event based video surveillance system with an open and extensible framework, *Machine Vision and Applications*, 19(5-6), 315–327, doi:10.1007/s00138-008-0153-z.

Vahdatikhaki, F. & Hammad, A. (2014), Framework for near real-time simulation of earthmoving projects using location tracking technologies, *Automation in Construction*, 42, 50–67, doi:10.1016/j.autcon.2014.02.018.

Vasenev, A.; Pradhananga, N.; Bijleveld, F.; Ionita, D.; Hartmann, T.; Teizer, J. & Dorée, A. (2014), An information fusion approach for filtering GNSS data sets collected during construction operations, *Advanced Engineering Informatics*, 28(4), 297–310, doi:10.1016/j.aei.2014.07.001.

Walsh, S. B.; Borello, D. J.; Guldur, B. & Hajjar, J. F. (2013), Data Processing of Point Clouds for Object Detection for Structural Engineering Applications, *Computer-Aided Civil and Infrastructure Engineering*, 28(7), 495–508, doi:10.1111/mice.12016.

Weerasinghe, T. I. P. & Ruwanpura, J. Y. (2010), Automated multiple objects tracking system ({AMOTS}), in *Proceedings of the Construction Research Congress*, 11–20.

Wu, C. (2011), VisualSFM: A Visual Structure from Motion System.

Yang, J.; Arif, O.; Vela, P. A.; Teizer, J. & Shi, Z. (2010), Tracking multiple workers on construction sites using video cameras, *Advanced Engineering Informatics*, 24(4), 428–434, doi:10.1016/j.aei.2010.06.008.

Yang, J.; Shib, Z. & Wua, Z. (2015), Automatic Recognition of Construction Worker Activities Using Dense Trajectories, in *Proceedings of the International Symposium on Automation and Robotics in Construction and Mining.*

Yang, J.; Vela, P. a.; Teizer, J. & Shi, Z. K. (2011), Vision-based crane tracking for understanding construction activity, *2011 ASCE International Workshop on Computing in Civil Engineering*, (February), 258–265, doi:10.1061/41182(416)32.

Zhai, D. & Goodrum, P. (2009), Relationship between automation and integration of construction information systems and labor productivity, *ASCE Journal of Construction Engineering Management*, 135(8), 746–753.

Zhang, C. & Elaksher, A. (2012), An unmanned aerial vehicle-based imaging system for 3D measurement of unpaved road surface distresses, *Computer-Aided Civil and Infrastructure Engineering*, 27(2), 118–129, doi:10.1111/j.1467-8667.2011.00727.x.