

**A VISUAL SENSING APPROACH TO ESTIMATE MATERIAL HAULING CYCLES IN HEAVY  
CONSTRUCTION AND SURFACE MINING JOBSITES**

\*E. Rezazadeh Azar<sup>1</sup>, and B.Y. McCabe<sup>2</sup>

<sup>1</sup>*University of Michigan  
Department of Civil and Environmental Engineering  
2350 Hayward Street  
Ann Arbor, MI, USA 48109  
(\*Corresponding author: ehsanra@umich.com)*

<sup>2</sup>*University of Toronto  
Department of Civil Engineering  
35 George Street  
Toronto, ON, Canada M5S 1A4*

---

# **A VISUAL SENSING APPROACH TO ESTIMATE MATERIAL HAULING CYCLES IN HEAVY CONSTRUCTION AND SURFACE MINING JOBSITES**

## **ABSTRACT**

Dump trucks are the main mean for hauling earth material within and outside of the construction and open-pit mining jobsites. The real-time monitoring of the material transportation trips is beneficial for proactive equipment management, volume estimation, and productivity measurement. The traditional manual observation is tedious, non-real-time, and expensive; therefore remote sensing technologies, namely GPS and RFID, have been employed for this reason. Ordinary construction videos are another cost-effective alternative to monitor dirt loading and hauling operations in real-time. A vision-based system has been developed to recognize and count hauling trips. This system has three modules including object detection, tracking, and action recognition. This system processes videos taken from access roads in which it recognizes dump trucks and then tracks them throughout the viewfinder. The action recognition module is designed to reject false alarms and also terminate tracking. This way, the system is able to count the number of hauling trips and also record the time between each trip. Several state of the art computer vision algorithms were employed to develop this system. Some of these algorithms were modified using spatiotemporal information to minimize the inherent errors of the vision-based methods. The experimental results showed considerable performance of the system for practical application in the construction and mining sectors.

## **KEYWORDS**

Computer vision, automated productivity measurement, hauling trip, dump truck, earth material.

## **INTRODUCTION**

Earthmoving operations in construction and mining jobsites, like other industrial activities, require real-time feedback for present and future purposes. Real-time equipment allocation and proactive safety monitoring are two instances of present applications, and the collected data can be used for cost estimation, simulation, and planning of future projects. Manual observation is costly, tedious, and error-prone; thus automated data collection systems have been applied to monitor earthmoving machines. More specifically, the permanent and repetitive nature of mining jobsites has encouraged faster adoption of new controlling technologies. For example, the application of automated computer-based dispatching systems in the mining industry dates back to late 1970's, when the communication and positioning tasks were limited to low-power VHF communications (Munirathinam & Yingling, 1994). GPS has been the main tool for tracking earthmoving fleets since the early 1990's to provide equipment locational data in determined time intervals. These spatiotemporal data can be interpreted using a logical reasoning or machine learning algorithms. Due to the similarity of operations and machines in the mining and heavy construction sectors, earthmoving was one the pioneers in adoption of automated data collection systems in construction (Navon, 2005). For instance, GPS-based monitoring systems have been employed to estimate the productivity of grading and leveling (Navon & Shpatnisky, 2005; Navon et al., 2004) and asphalt paving (Navon & Shpatnisky, 2005) operations. These productivity measurement systems map the locational data into the work zones, and then interpret the movements of the machines to estimate their productivity. In addition, the application of more than one GPS receiver on two sides of grader and bulldozer blades can facilitate achieving accurate excavation and earth-fill profiles.

---

Material transportation makes up 50% to 60% of the cost in open-pit mining operations, which is mainly done using large-scale shovels, excavators and dump trucks (Alarie & Gamache, 2002). Earth material hauling also represents a costly portion of heavy construction projects, depending on a project's type and conditions. For instance, it consists the main activity of earth-fill dams, railways, and highway construction projects. As a result, several research projects as well as a number of commercial products have been developed for controlling material hauling fleets. For instance, loading and dumping zones were manually defined and the spatiotemporal data of dump trucks and a loader were beacons on a map of the jobsite. Then software estimates the cycle times based on the stop and travel times between the loading and dumping zones (Hildreth et al., 2005). Other radio-based sensing device, radio-frequency identification (RFID), was also tested for estimating earthmoving productivity. RFID readers can register the proximity of tags; their effective range depends on whether the tags are active (longer range) or passive (shorter range). These readers were installed on the entrance gates of loading and dumping zones to register entrance and exit times of the tagged trucks. The cycle times can be estimated using these records (Montaser & Moselhi, 2012). Both GPS and RFID systems are active monitoring methods which require installing tags on every machine. The installation process requires manual preparation, which can be problematic for rental equipment. Moreover, the provided spatiotemporal data are limited to time and location, which cannot accurately represent machine actions.

Generally, surface earthmoving sites are open fields where relatively clear sightlines can be selected. The emergence of low-cost digital cameras and high-capacity storage devices has encouraged the use of time-lapse and video cameras in these sites. Moreover, vision-based controlling systems are not restricted by manual tagging of individual machines and can monitor operations without any interruption. These factors offer an appropriate opportunity to use vision-based data collection systems for recognition and estimation of earthmoving cycles. These state of the art algorithms were recently used for productivity estimation of excavation (Gong & Caldas, 2011) and dirt loading cycles (Rezazadeh Azar et al., 2012; Heydarian, et al. 2012).

This paper introduces a vision-based system to monitor one of the main earthmoving operations, earth material hauling, which is common in heavy construction and mining sectors. First it explains the modules of this system, and then it describes the architecture of the framework to recognize and count hauling trips. Next, the performance of the system is evaluated using test videos from two rock-fill dam construction projects. Finally, the shortcomings and potential applications of this prototype software are discussed.

## **MODULES OF THE SYSTEM**

### **Object Recognition**

The object recognition module is responsible for detecting off-highway dump trucks as the primary indicator of hauling operations in mining and heavy construction jobsites. In general, there are two main approaches to detect objects in videos. The first method is to extract static frames from the video stream and search for the objects of interest using visual features such as edges and shapes. Algorithms in the second method use dynamic features, such as optical flow or foreground-background segmentation, to isolate moving objects and then identify them. This approach, however, is sensitive to high amounts of visual noise and occlusions caused by moving objects, e.g. two object move close together. Earthmoving jobsites are harsh environments in which there are considerable amounts of dynamic visual noise, such as dust and equipment exhaust, and machines moving continuously; therefore the first approach was selected for this research project. Experimental results showed that the Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005) performs well in detection of dump trucks using eight viewpoints (Rezazadeh Azar & McCabe, 2012). Figure 1 shows these viewpoints. The main identified issue was the computation runtime which took about 26 seconds to search for eight orientations in a regular VGA frame (640 x 480 pixels). This problem was resolved using parallel computing on a multicore graphics processing unit (GPU) which made the same process possible in about one second (Rezazadeh Azar et al., 2012). Thus, the system is able to maintain real-time video stream by scanning frames in time intervals slightly farther apart

---

than one second. Due to road conditions, heavy loads, and safety reasons, dump trucks move slowly in jobsites (<25 km/h) and therefore, very short time intervals, e.g. less than a second, are not necessary to guarantee catching a passing dump truck in at least one scanned frame. Four second intervals were recommended to study equipment movement in jobsites (Rojas, 2008) which is also used in this research.



Figure 1 - Left: orientations; Right: examples of views (from top and left to right: front, front-left, front-right, side-left, side-right, rear, rear-left, rear-right)

All of the existing object recognition algorithms, including HOG method, are prone to two types of error: false positives and false negatives. As it is described later in this paper, an action recognition module is developed to reject false alarms. Despite lowering object recognition thresholds, the possibility of missing objects, however, still remains a challenge.

## Tracking

Tracking is a very diverse and evolving field of computer vision. There are several approaches to track an object of interest, such as kernel-based, contour-based, and probabilistic methods. Recent studies showed that the mean-shift method could provide promising results in the tracking of construction resources (Park et al., 2011; Gong & Caldas, 2011). Mean-shift is a kernel-based algorithm that theoretically can track the visual features of an object that can be translated to a weighted dataset. Color histograms and edges are two instances of these visual features. The mean-shift method, however, is sensitive to the proximity of objects with similar features, resulting in the expansion of the tracker frame to capture both objects or switching from the main target to the nearby one.

A hybrid tracking algorithm was developed that performs well in tracking slow-moving objects, such as dump trucks (Rezazadeh Azar et al., 2012). This tracking method consists of the HOG object recognition and KLT feature tracking (Tomasi & Kanade, 1991). In this framework, the KLT method tracks the center of the identified object for a short specified time, e.g. one second, and then the prior bounding box of the tracking object is projected to the new frame with the new center. At the same time, the HOG object recognition part searches the new frame for the object of interest. The recognition thresholds are set low to minimize the risk of false negatives which also increases the number of false positives. The projected box has two main applications. First it is used to select the target among detected boxes, and secondly, it eliminates the risk of false negatives in each search trial as at least one box exists in every frame. In each search frame, there will be a projected box and possibly a detected target plus other true positives and false alarms. A simple disjoint-set data structure algorithm is used to group the detected rectangle that is spatially coherent with the projected box, producing a new bounding box and removing other detections. This method considers two boxes in the same subset if their bounding areas overlap. All the distances between x and y elements of the matching corners of two boxes should be less than the minimum average of the height and width of the rectangles times a threshold to merge two boxes (Viola & Jones, 2001). The corners of the merged rectangle are the average of the corners of the projected box and the overlapping detection. If none of the detections is temporally coherent, the projected rectangle will be set as the final box.

The slow movement of earthmoving machines in videos enabled us to optimize this recognition process for runtime efficiency in two ways. First, the recognition module scans the frame for three orientations instead of eight, including the prior orientation and the two adjacent viewpoints, which reduces the recognition process from 1 to 0.39 seconds. For example, if a dump truck has front-right orientation, the recognition part of the hybrid tracker will only search for front, front-right, and side-right orientations. This does not compromise the results due to the fact that the dump trucks cannot drastically change their orientations in a short time, e.g. one or two seconds. Second, the HOG recognition engine searches for the three viewpoints in a region of interest (ROI) that is determined using the prior location and size of the target. The center of this ROI is the same center as the bounding box of the target in prior search frame; but the width and height of the ROI are two and half times larger than the target's bounding box. This way, the ROI is dynamically adjusted for each trial, allowing the system capture movement and scale changes of the tracking machine.

The dynamic structure of this hybrid tracking algorithm allows handling scale and orientation changes of a slow moving target. This is achieved with the merger of the detected object with the projected box, thereby adjusting the size of the box to handle scale changes of the target. Also, the three search orientations are reset in each search trial based on the prior detection. This way, it dynamically changes search orientations and therefore can keep track of orientation changes of the target.

### **Action Recognition and System Architecture**

The action recognition module is responsible for rejecting false positive detections, keeping track of true positive dump trucks, and counting the number of hauling trips in each direction. In the first stage, the object recognition module searches for dump trucks in eight orientations in a predetermined time interval, 3 and 4 seconds for this experiment. The thresholds of the object recognition module are set rather low to avoid false negatives, even though it may result in more false positives. For example, both true positive and false alarm are detected in frame b in Figure 2. False negatives are costlier than false positives as spatiotemporal information of the video and the domain knowledge can be used to remove false alarms, but false negatives cannot be offset. Then all detected boxes are passed to the tracking module. This module tracks all the boxes regardless of being true or false positive.

The tracking module provides useful spatiotemporal data that can be applied to distinguish true positives from false alarms. False positives occur in two ways: random scattered boxes, and repetitive detections with the same size and location. The action recognition module benefits from the recognition part of the hybrid tracking method to remove the first group of false alarms. As explained before, the hybrid tracker uses consecutive recognitions of dump trucks to track the target. The system assigns a zero score to all detections before passing them to tracking module. Any successive detection of the targets in the following tracking time intervals, one second for this framework, increases their score by one. As a result, a tracking object can have 0-4 score in a 4 second and 0-3 in a 3 second interval. The system eliminates tracking targets with scores less than 2. Since random false positives do not reoccur, this constraint is able to remove them. For instance, this reasoning algorithm successfully removed the false positive in frame c in Figure 2.

In the second group, the recognition part of the hybrid tracker repeatedly detects a false window in the same place. Thus, another constraint was developed which rejects a set of consecutive tracking boxes without 5% displacement or scale change. The scale part is defined for the cases where dump trucks move aligned with the camera viewpoint (e.g. frame right in Figure 4). The tracking boxes of these machines do not displace; however, their sizes change depending whether they are moving toward camera or going away. This controller may eliminate a motionless true positive (a temporarily unmoving dump truck on the road), but this does not affect the outcome as the system will detect and track that machine upon departure.

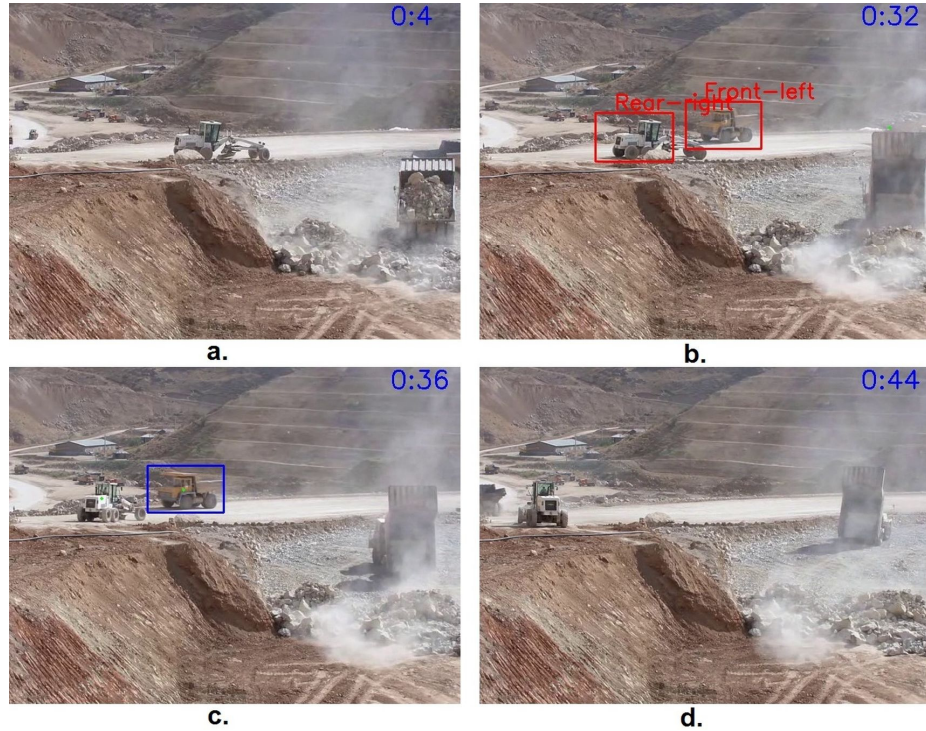


Figure 2 - frame a: initial view of the scene; frame b: one of the object recognition trials, including a true positive and a false alarm; frame c: tracking result after 4 seconds; frame d: termination of tracking

Defining the tracking termination criteria is the next problem. The exit of a tracking dump truck from the frame should terminate tracking (e.g. frame d in Figure 2). In some views, however, dump trucks disappear in the video without going out of the frame. For instance, they may go behind a hill or move aligned with the camera view and become smaller than the smallest search window, 104x96 for front and rear views and 1280x80 pixels for the rest of orientations, in which they become undetectable (e.g. frame right in Figure 4). Thus, the second termination constraint finishes the tracking of a confirmed target if it cannot get more scores in the following intervals. Figure 3 depicts the entire flowchart of the system in which different modules of the system are highlighted by color.

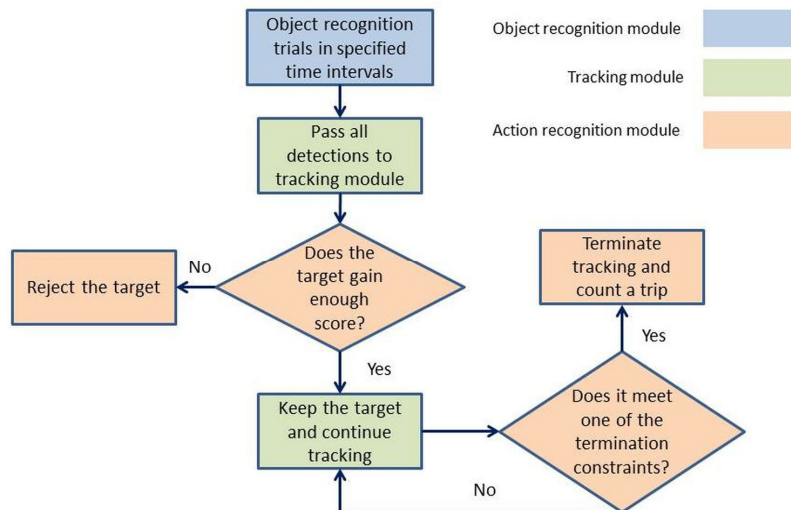


Figure 3 - Flowchart of the system



## EXPERIMENTAL RESULTS

Three test videos with total duration of 23 minutes were used to evaluate the performance of the system. These videos were taken from two rock-fill dam construction sites where the off-highway dump trucks carry rock-fill material from quarry to dam body. One of the authors took two videos from slightly different viewpoints in one of the sites in which both show a side view of an access road to quarry (frame left in Figure 4). The third video was recorded by a person in another jobsite who was in arm's length of this research project. He was asked to record a video using a stable camera showing dump trucks hauling earth material. This video was captured from an overlooking viewpoint aligned with the road's axis (frame right in Figure 4). Altogether, 23 dump trucks passed the views. In addition to off-highway dump trucks, at least eight other types of construction equipment, such as grader, loader, and pickup truck, appeared in the videos.



Figure 4 - Some of the material transportation views

The system processed these videos with two different object recognition time intervals, 3 and 4 seconds. The results are presented in Table 1. The first column provides types of data, including manual observation and machine-generated data using 3 and 4 second object recognition intervals. The second column presents the number of correctly recorded trips. The system could not identify 2 and 3 trips in 3 and 4 second intervals, respectively. Despite these false negatives, the action recognition module successfully rejected all of the false positives. The fifth columns depicts the trips that correctly identified by the system, but a foreground object fully masked the target which resulted in missing the target. The system recognized the reappeared machine, resulting in double registration. The average time difference between the machine-generated and manual observation are presented in the last column. The termination point of each case (exit from the frame or become small enough) was compared with actual location of the machine to measure the difference. For example, the system lost track of a machine for a few seconds before it actually departed the frame. This data is useful for accuracy assessment of the system to measure the hauling cycle times.

Table 1 - Results of the experiments with 3 and 4 second recognition intervals on test videos

Data type	Number of detected trips	False negatives	False positives	Incomplete detected cycles	Average difference with ground truth (Sec)
Ground truth	23	-	-	-	-
Machine-generated 3 Second interval	21	2	0	2	2.1
Machine-generated 4 Second interval	20	3	0	2	1.7

## DISCUSSION

The application of remote video camera systems has become popular in construction and mining sectors for real-time monitoring of the jobsites. Numerous companies provide specialized live streaming cameras for rugged construction sites at relatively low-cost. A real-time computer vision system has been developed for automated interpretation of the videos showing material transportation which can be processed on a regular computer. The cycle recognition accuracy rates were 91% and 87% in 3 and 4 second intervals, respectively. This system, however, showed two major shortcomings. First it was not able to detect 9-13% of the machines that experienced major visual noise (e.g. frame a. in Figure 5). Second, the system showed difficulty in tracking occluded machines, because the foreground object completely distracts the feature tracker, so it cannot correctly project the box. More importantly, the recognition part of the hybrid tracker has no sightline of the target, and therefore is not able to correct the tracking process. This issue was only observed in the ground views where the foreground machines fully obstructed the target (frame b. in Figure 5). The action recognition module could quickly identify this problem through a low tracking score, which terminates tracking. This would not harm the results if the main target exits the frame meanwhile. But if it remains in the camera view, the object detection may redetect the machine which can result in extra trip count. This issue can be resolved using spatiotemporal reasoning which was not investigated in this research. Elevating the camera view is another resolution for this problem. Elevated cameras have semi-occluded view which the hybrid tracker could successfully track the object of interest (Rezazadeh Azar et al., 2012).



Figure 5 - frame a: inability to detect due to visual noise; frame b: occlusion misleads the tracker

Proper location of the camera has a key role in providing correct results. Elevated views are best for installing the cameras as passing machines do not completely mask each other. In addition, the effect of observed visual noise, such as dust, will be slightly lower from an above ground view. The pitch angle of the elevated viewpoints, however, should not very steep as the dump trucks may become difficult to detect.

## CONCLUSION

A real-time controlling system was developed to recognize and count the number of hauling trips in open-pit mining and construction jobsites. This system uses HOG recognition algorithm to detect candidates and then passes them to a novel hybrid tracking framework. This hybrid tracking algorithm uses an optimized combination of HOG recognition and a classic feature tracking method to track dump trucks. Action recognition is the third module of this system which uses the scores given by the recognition part of the hybrid tracker to reject false detections. Exiting from the frame and becoming undetectable are two main factors to terminate tracking and record a trip. The experimental results on three videos from actual hauling operations provided promising results in detection and estimation of earth material hauling cycles. Appropriate camera viewpoints, including elevated viewpoints, are key factors in successful detection and tracking of dump trucks. Future work will focus on the recognition of individual machines for productivity estimation.



## REFERENCES

- Alarie, S., & Gamache, M. (2002). Overview of Solution Strategies Used in Truck Dispatching Systems for Open Pit Mines. *International Journal of Surface Mining, Reclamation and Environment*, 16(1), 59–76.
- Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *Proc. Computer Vision and Pattern Recognition (CVPR 2005)*, (1, pp. 886 – 893).
- 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.
- Heydarian, A., Golparvar-Fard, M., & Niebles, J.C. (2012). Automated Visual Recognition of Construction Equipment Actions Using Spatio-Temporal Features and Multiple Binary Support Vector Machines. *2012 Construction Research Congress*, (pp. 889-898), West Lafayette, IN.
- Hildreth, J., Vorster, M., & Martinez, J. (2005). Reduction of Short-Interval GPS Data for Construction Operations Analysis. *Journal of Construction Engineering and Management*, 131(8), 920–927.
- Montaser, A., & Moselhi, O. (2012). RFID+ for Tracking Earthmoving Operations. *2012 Construction Research Congress*, (pp. 1011-1020), West Lafayette, IN.
- Munirathinam, M., & Yingling, J.C. (1994). A review of computer-based truck dispatching strategies for surface mining operations. *International Journal of Surface Mining, Reclamation and Environment*, 8(1), 1-15.
- Navon, R. (2005). Automated project performance control of construction projects. *Automation in Construction*, 14(4), 467– 476.
- Navon, R., Goldschmidt, E., & Shpatnisky, Y. (2004). A concept proving prototype of automated earthmoving control. *Automation in Construction*, 13(2), 225– 239.
- Navon, R., & Shpatnisky, Y. (2005). Field Experiments in Automated Monitoring of Road Construction. *Journal of Construction Engineering and Management*, 131(4), 487– 493.
- Park, M.-W., Makhmalbaf, A., & Brilakis, I. (2011). Comparative study of vision tracking methods for tracking of construction site resources. *Automation in Construction*, 20(7), 905-915.
- Rezazadeh Azar, E. & McCabe, B. (2012). Automated Visual Recognition of Dump Trucks in Construction Videos. *Journal of Computing in Civil Engineering*, 26(6), 769–781.
- Rezazadeh Azar, E., Dickinson, S., & McCabe, B. (2012) Server-Customer Interaction Tracker; A Computer Vision-Based System to Estimate Dirt Loading Cycles. *Journal of Construction Engineering and Management*, in press.
- Rojas, E.M. (2008). *Construction Productivity: A Practical Guide for Building and Electrical Contractors*. J. Ross Publishing, Fort Lauderdale, FL.
- Tomasi, C., & Kanade, T. (1991). *Detection and tracking of point features*. (Technical Report No. CMU-CS-91-132). Carnegie Mellon University, Pittsburgh, PA.
- Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proc. Computer Vision and Pattern Recognition (CVPR 2001)*, (1, pp. 1 – 9).
-