A computer vision and reasoning based video interpretation system for rapid productivity analysis of construction operations

Jie Gong
Dept. of Construction, Southern Illinois University Edwardsville, IL 62025, USA
Carlos H. Caldas
Dept. of Civil, Architectural & Environmental Engineering, University of Texas at Austin, USA

Abstract

On construction jobsites, considerable manual data collection and analyses efforts are often required to conduct productivity studies. Intelligent computer vision systems that can autonomously observe the behavior of construction objects and infer their working status have the potential to enable automated productivity assessment and therefore reduce some of the shortcomings of the manual methods. As an important step towards developing field deployable intelligent vision systems, this research addresses the problem of automated productivity data collection through the development of an intelligent construction video interpretation methodology and the implementation of an intelligent video computing system based on the developed methodology. The proposed methodology was validated through using the developed prototype system to analyze five construction video sequences that record various types of construction operations. The developed system was able to interpret these videos into productivity information with an accuracy that was close to manual analysis, without the limitations of onsite human observation.

Keywords: productivity, computer vision, video interpretation, automated data collection and analysis

1 Introduction

Productivity information is the blood running through the veins of construction production systems. If construction activities can be viewed as a flow of material, labor, tool, and equipment at the workface, accurate and timely productivity information is the key to ensure the efficient flow of construction resources and to minimize waste. At the same time, however, gathering productivity information in construction projects is a difficult task. Construction sites are dynamic and complex. Site productivity data gathering methods have been mostly manual and have shown little sign of improvement over years (Cheok et al. 2000).

Recently, considerable research in the construction domain has been centered on sensing systems for automated onsite productivity data collection. These studies fall in two broad areas: (1) automated project progress tracking for measuring construction output (Cheok et al. 2000; Bosche and Haas 2008); and (2) automated resource utilization tracking for measuring construction input (Su and Liu 2007; Navon and Sacks 2007; Zou and Kim 2007; Kim and Bai 2009). In essence, automated resource utilization tracking focuses on equipment hours and labor hours, since these are the two most common types of resources measured to assess productivity. In project control systems, equipment hours and labor hours are measured from a holistic point of view: both effective hours and non-effective hours are rolled into final work hours for project controls. To go beyond this level of detail, video
monitoring methods can be used to automatically measure how these work hours have been spent, whether productively or not.

The objective of this research is to develop a video interpretation methodology for automated productivity analysis of construction operations. More specifically, this research focuses on methods of modeling and automating the workflow in video-based productivity analysis, methods of leveraging existing knowledge context to support information extraction, and computational approaches toward observing and reasoning about construction resources and their interactions.

2 A Common Workflow in Video-based Construction Productivity Analysis

A common process of video-based productivity data collection essentially involves four parties, including a data collector, working crew and foreman, engineers, and managers at certain levels, and four sequential sub-processes, which are preparation, video recording, review and analysis, and implementation (Oglesby et al. 1989). A graphical representation of this process is shown in Figure 1. Notably, the majority of data collection efforts are taken by the data collector. A data collector typically makes the following preparations prior to actual recording: (1) plan the camera position; and (2) record relevant information to aid understanding during subsequent viewing. After videos have been recorded, a data collector can choose to conduct informal review with the workforce to brainstorm improvement methods. Then the findings are presented to managers to seek approval for implementation. In addition to this, a formal review and analysis session can be held with all involved parties to seek improvement methods.

Moreover, regardless of the type of data analysis approaches implemented, the reality is that intensive manual effort is pervasive in such a process, particularly in the tasks of “informal review”, “present findings”, and “formal analysis and review” (Su and Liu 2007; Kim and Bai 2009). The delay and expense caused by intensive manual efforts can extend beyond individual tasks across the whole spectrum of the process map, causing high data collection expenses and the dilution of the effectiveness of such a program. Consequently, jobsite leadership frequently considers construction cameras as just a contractual tool for site surveillance, dispute resolution, and site demonstration.

3 A Computational Representation of Video-based Productivity Analysis

To define a computational representation of the process map as shown in Figure 1, two steps are involved. The first step is to replace the manual steps and processes in video analysis with computational methods. The second step is to model the mechanism of human reasoning in construction video analysis for providing computers with a reasoning logic that imitates that of humans. A detailed explanation of the development can be found in Gong and Caldas (2009).

For the first step, a variety of computational methods, typically in the domain of computer vision, can be found to automate or support the processes in video-based construction productivity analysis. It
is reasonable to expect that there are four essential steps in manual construction video analysis after a construction video is recorded. They are: (1) recognize what objects are in the video; (2) understand what is happening in the video; (3) summarize what happened in the video; and (4) review what happened in the video. Correspondingly, four computing methods, which include visual recognition and tracking, model-based reasoning, video content organization, and video retrieval, can be used to represent these essential manual analysis steps. Among them, the model-based reasoning is a three-step reasoning process that represents a hierarchical reasoning procedure. These steps are state classification, event detection, and scenario recognition. They are proposed to recover three principle different types of information, be they resource utilization, work flow, and inefficiency. More specifically, resource utilization concerned with the time utilized in different working states by construction resources. It is desirable logically to have construction resources spend more time in productive states. Work flow is essentially about working sequences. And inefficiency is rooted in abnormal production scenarios.

The second step involves the design of a structured reasoning mechanism. In other words, these elements and processes described above should be able to recognize a set of rules, heuristics, and logics that they can use to automatically carry out the information extraction tasks. Human beings usually reasons within a context. More precisely, when human analyzes construction videos for the purpose of productivity analysis, they utilize their knowledge and experiences as background information when they process the visual contents in the video. In the tradition process as shown in Figure 1, a data collector usually uses line sketches, notes, or even mental memories as means of recording contextual information. It is how to represent and describe a similar context for computers that this reasoning mechanism should address. Specifically in this research, the reasoning mechanism is designed to support visual recognition and tracking and video reasoning. The needs of information support for these two elements and the proposed knowledge representations that can be used to address these needs are summarized in Table 1.

<table>
<thead>
<tr>
<th>Elements in Work Flow</th>
<th>Need of Information Support</th>
<th>Supporting Knowledge Representations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Recognition and Tracking</td>
<td>Which objects should be recognized and tracked?</td>
<td>The concepts of production unit and the method’s leading resources in Method Productivity Delay Model</td>
</tr>
<tr>
<td></td>
<td>What information is needed and available to drive the video reasoning process?</td>
<td>Spatial context</td>
</tr>
<tr>
<td>Model-based Video Reasoning</td>
<td>Semantic context</td>
<td>Task elements</td>
</tr>
</tbody>
</table>

To explain the usage of either process models or process charts to support video interpretation, an example of slab pour operation is shown in Figure 2. The process of this slab pour operation consists of a number of steps. These steps include loading the bucket, bucket ready, pouring the concrete, and bucket departure. A process model is developed to represent these steps in a concise way (Figure 2). First, every state in the process model stands for a type of working status, determining the categories of the working states that need to be classified in video processing (semantic context). Second, the location description in the process model can be used to divide the image scene into various work zones through image surface marking (spatial context), each of them corresponding to a unique working state. In addition, the sequence of operations as defined in the earthmoving process model determined the scheme of state transition (temporal context), constituting the temporal context for
detecting any sequence violation, which is typically a sign of off-track operation. Finally, each of those working states can be assigned within a time constraint that is simply the maximum amount of time allowed to be spent within that specific state during each cycle.

Finally, the elements and reasoning mechanisms that are resulted from the development in the step one and the step two can be collectively assembled into a global computational representation. This representation exhibits the overall structure of the video interpretation methodology (Figure 3).

4 Development of Construction Object Recognition and Tracking Methods

Each of the conceptual computational steps in the video interpretation methodology requires a set of supporting algorithms. This section describes the development of a test bed that supports the development and integration of computer vision algorithms as well as the characterization of the performance of the developed algorithms in construction environments. These algorithms can be used to observe and track the objects of interest in recorded construction videos. The essential architecture of the proposed test bed is a video processing pipeline that includes 6 modules (Figure 4a). These modules are object recognition, new blob entrance detection, blob tracking, trajectory post-processing, trajectory generation, and motion analysis. Herein, a blob refers to a connected region in an image, which is typically so designated because it shows some degree of similarity in certain measures and is likely to correspond to an object. The interfaces between these modules have been standardized, allowing each module to be changed without impacting other modules. The following sections briefly describe the recognition and tracking methods that were developed and tested within this video processing framework.

The algorithms that were developed in the module of object recognition include a colour-based recognition method, a method based on the cascade of simple features, and three foreground/background segmentation methods including Gaussian Mixture Models, Codebook-based methods, and Bayesian Models. These algorithms were also tested on a common set of construction videos to characterize their performances. It was found through the evaluation that Colour-based and Viola Jones-based recognition methods provide a general approach to train models for specific construction objects. However, these two general methods cannot be directly applied to recognize objects. A large number of photo samples are needed to train specific models for different objects. But, once the models are trained, the performances of these models are robust and fast. Background subtraction-based methods can be directly applied on construction videos to isolate interested objects from the background scenes. In particular, it was found that the Codebook-based and the Bayesian-based methods are more suitable than the Gaussian Mixture-based method to be used in construction scenarios because of their ability to handle sudden changes. But, the downside of
foreground/background segmentation methods is that background subtraction tends to identify all foreground objects without differentiating which foreground objects are objects of interests. This may work in controlled construction scenarios where unexpected construction objects are not frequently present.

Two general types of tracking methods, including Filtering and Data Association and Target Representation and Localization, were implemented and tested in the module of “Blob Tracking”. Filtering and Data Association is a top-down process that deals with the dynamics of the tracked object, awareness of scene priors, and evaluations of different hypotheses. Typical example algorithms are Kalman Filter and Particle Filter. Target Representation and Localization is a bottom-up process to cope with changes in the appearance of the target. A typical method in this category is Mean Shift. In practices, the methods mentioned above are often used in combination. This gives rise to six different tracking methods that were incorporated into the “Blob Tracking” module. The performances of these tracking methods were tested on videos with different complexities. Based on the performance evaluation, the findings can be summarized as: (1) Mean Shift method outperforms other methods such as Kalman filters and particle filters; and (2) combinations of mean shift methods with Kalman Filters and Particle Filters are highly recommended, since this kind of combinations exacts the best innate characteristics from each method.

The test bed was written using C++, and it integrates several commercial strength libraries, including Microsoft Directshow SDK (System Development Kit), Microsoft Foundation Class, Intel Open Source Computer Vision Library, and XD++ Flowchart SDK. The graphic interface of the test bed is shown in Figure 4b.

![Figure 4. (a) The System Diagram for Video Processing; (b) The Interface of the Test Bed](image)

5 Case Studies

Five video sequences, which record five different types of construction operations, were analyzed using the developed video interpretation system to demonstrate the applicability of the developed methodology in different real construction operations as well as to assess their performance based on the accuracy and speed of productivity information extraction. The analyzed operations include three cyclic operations (earthmoving, slab pour, and column pour) and two non-cyclic operations (scaffold installation and material hoisting). The analysis results of these videos are summarized in Table 2. It can be noted that the average accuracy of the proposed methodology is around 87%. Moreover, the proposed methodology can interpret construction video contents at a rate of twenty frames per second given the properly defined contextual information. Therefore, the interpretation can happen in near real-time. In other words, if a real-time video stream is fed into this system, it interprets the video instantaneously and delivers updates of operation state, event, and scenario. Conversely, it is expected that in the traditional human review based approach one must look at the film three to five times
before the interference, wasted motion, lost time, duplicated effort, and a myriad of other factors
become apparent (Oglesby 1989).

Table 2 Summary of Analysis Results for Example Video Analysis

<table>
<thead>
<tr>
<th>Analyzed Operations</th>
<th>Knowledge Representation</th>
<th>Object Recognition Method</th>
<th>Object Tracking Method</th>
<th>Accuracy (Time Utilization)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column Pour (70 Minutes)</td>
<td>Process Model</td>
<td>Cascade of Simple Features</td>
<td>N/A</td>
<td>90%</td>
</tr>
<tr>
<td>Slab Pour (28 Minutes)</td>
<td>Process Model</td>
<td>Cascade of Simple Features</td>
<td>N/A</td>
<td>86%</td>
</tr>
<tr>
<td>Earthmoving (12 Minutes)</td>
<td>Process Model</td>
<td>Bayesian Background Model</td>
<td>Kalman Filter + Mean Shift</td>
<td>90.6%</td>
</tr>
<tr>
<td>Scaffold Installation (14 Minutes)</td>
<td>Process Charts</td>
<td>Bayesian Background Model</td>
<td>Kalman Filter + Mean Shift</td>
<td>81.2%</td>
</tr>
<tr>
<td>Hoisting (6 Minutes)</td>
<td>Process Charts</td>
<td>Color-based recognition model</td>
<td>Kalman Filter + Mean Shift</td>
<td>86.7%</td>
</tr>
</tbody>
</table>

6 Conclusions

The proposed methodology reshapes the original workflow that is utilized in the traditional video
review based methods. The proposed elements and processes as well as video reasoning mechanism in
the developed methodology enable analyzing construction operation systematically in a coherent way.
The resulted new workflow presents significant advantages over the traditional workflow. A direct
benefit with the proposed methodology is the support for rapid onsite method correction and the
availability of organized video contents after the video interpretation. The immediate available
information about the production status of ongoing construction operations empower the onsite
management team to quickly respond to inefficiency that may be caused by improper working
methods. Also, as a whole, the new workflow can potentially enable the collection of productivity
information regarding ongoing construction operations remotely from offsite. Therefore, it saves the
travelling time and cost that would be incurred by sending dedicated personnel to different sites to
conduct video-based operation analysis.

References

*Automation in Construction*, 17(4), 499-512.
HILDRETH, J., ASCE, M., VORSTER, M., and MARTINEZ, J., 2005. Reduction of Short-Interval GPS Data for
*Automation in Construction*, 16, 474-484.
New York.