Advanced Engineering Informatics 27 (2013) 652–663

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers

Mani Golparvar-Fard [a,*], Arsalan Heydarian [b,1], Juan Carlos Niebles [c,2]

[a] Department of Civil and Environmental Engineering, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States
[b] Charles E. Via Department of Civil & Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, United States
[c] Department of Electrical and Electronic Engineering, Universidad del Norte, Barranquilla, Colombia

ARTICLE INFO

Article history:
Received 19 May 2012
Received in revised form 26 July 2013
Accepted 9 September 2013
Available online 13 October 2013
Handled by W.O. O'Brien

Keywords:
Computer vision
Action recognition
Construction productivity
Activity analysis
Time-studies
Operational efficiency

ABSTRACT

Video recordings of earthmoving construction operations provide understandable data that can be used for benchmarking and analyzing their performance. These recordings further support project managers to take corrective actions on performance deviations and in turn improve operational efficiency. Despite these benefits, manual stopwatch studies of previously recorded videos can be labor-intensive, may suffer from biases of the observers, and are impractical after substantial period of observations. This paper presents a new computer vision based algorithm for recognizing single actions of earthmoving construction equipment. This is particularly a challenging task as equipment can be partially occluded in site video streams and usually come in wide variety of sizes and appearances. The scale and pose of the equipment actions can also significantly vary based on the camera configurations. In the proposed method, a video is initially represented as a collection of spatio-temporal visual features by extracting space–time interest points and describing each feature with a Histogram of Oriented Gradients (HOG). The algorithm automatically learns the distributions of the spatio-temporal features and action categories using a multi-class Support Vector Machine (SVM) classifier. This strategy handles noisy feature points arisen from typical dynamic backgrounds. Given a video sequence captured from a fixed camera, the multi-class SVM classifier recognizes and localizes equipment actions. For the purpose of evaluation, a new video dataset is introduced which contains 859 sequences from excavator and truck actions. This dataset contains large variations of equipment pose and scale, and has varied backgrounds and levels of occlusion. The experimental results with average accuracies of 86.33% and 98.33% show that our supervised method outperforms previous algorithms for excavator and truck action recognition. The results hold the promise for applicability of the proposed method for construction activity analysis.

1. Introduction

Equipment activity analysis, the continuous process of benchmarking, monitoring, and improving the proportion of time construction equipment spend on different construction activities, can play an important role in improving construction productivity. A combination of detailed assessment and continuous improvement can help minimize idle times, improves operational efficiency [1–5], saves time and money [6], and results in a reduction of fuel use and emissions for construction operations [7,8]. Through systematic implementation and reassessment, activity analysis can also extend equipment engine life and provide safer environments for equipment operators and workers.

Despite the benefits of activity analysis in identifying areas for improvement, an accurate and detailed assessment of work in-progress requires an observer to record and analyze the entire equipment's actions for every construction operation. Such manual tasks can be time-consuming, expensive, and prone to errors. In addition, due to the intra-class variability on how construction tasks are typically carried out, or in the duration of each work step, it is often necessary to record several cycles of operations to develop a comprehensive analysis of operational efficiency. Not only the traditional time-studies are labor intensive, but they also require a significant amount of time to be spent on manually analyzing data. The monotonous data analysis process can also affect the quality of the process as a result of the physical limitations or biases of the observer. Without a detailed activity analysis, it is unfeasible to investigate the relationship between the activity duty cycles vs. productivity, or fuel use and emissions [9]. There is a need for a low-cost, reliable, and automated method for activity analysis that can be widely applied across all construction projects.

* Corresponding author. Tel.: +1 217 300 5226; fax: +1 217 265 8039.
E-mail addresses: mgolpar@illinois.edu (M. Golparvar-Fard), aheydar@vt.edu (A. Heydarian), njuan@uninorte.edu.co (J.C. Niebles).
[1] Tel.: +1 (540) 383 6422; fax: +1 (540) 231 7532.
[2] Tel.: +57 (5) 350 9270; fax: +57 (540) 231 7532.

1474-0346/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aei.2013.09.001

Fig. 1. Example frames from video sequences in excavator and truck action video dataset. Excavators: (a) digging; (b) hauling (swinging with bucket full); (c) dumping; and (d) swinging (bucket empty); Trucks: (e) filling; (f) moving; and (g) dumping.

This method either vision-based or non-vision based, needs to *remotely* and *continuously* analyze equipment's actions and provide detailed field data on their performance.

Over the past few years, cheap and high-resolution video cameras, extensive data storage capacities, and the availability of Internet connection on construction sites have enabled capturing and streaming construction videos on a truly massive scale. Detailed and dependent video streams provide a transformative potential for gradually and inexpensively sensing actions of construction equipment, enabling construction companies to remotely analyze operational details and in turn assess productivity, emissions, and safety of their operations [10]. To date, the application of existing site video streams for automated performance assessment is still untapped and unexploited by researchers in most parts.

Here, we address a key challenge: *action recognition*; i.e., determining various actions equipment performs over time. While in the past few years several studies have looked into these areas (Section 2), many challenging problems still remain unsolved. As a step forward, this paper focuses on the problem of recognizing single actions of earthmoving equipment from site video streams. Fig. 1 shows examples of the actions of an excavator and a dump truck operation, wherein the excavator performs a cycle of digging, hauling (swinging with a full bucket), dumping, and swinging (with an empty bucket) and the truck performs a cycle of filling, moving, and dumping.

Given videos taken by a fixed camera with small lateral movements (caused by wind or small ground vibrations), clutter, and moving equipment, the task is to automatically and reliably identify and categorize such actions. This paper presents an algorithm that aims to account for these scenarios. As such, the state-of-art research in this area is first overviewed. Next, a set of open research problems for the field are discussed, including action recognition under different camera viewpoints within dynamic construction sites. The specific focus of the proposed method and its details are then described. Also, a comprehensive dataset and a set of validation methods that can be used in the field for development and benchmarking of future algorithms are provided. The perceived benefits and limitations of the proposed method in the form of open research challenges are presented.

2. Background and Related Work

In most state-of-the-art practices, the collection and analysis of the site performance data are not yet automated. The significant amount of information required to be manually collected may (1) adversely affect the quality of the analysis, resulting in subjective reports [11,12] and (2) minimize opportunities for continuous monitoring which is a necessary step for performance improvement [11–14]. Hence, many critical decisions may be made based on this inaccurate or incomplete information, ultimately leading to project delays and cost overruns.

In recent years, a number of research groups have focused on developing techniques to automatically assess construction performance. The main goal of these methods is to support improvement of operational efficiency and minimize idle times. Several studies such as [1–4] emphasize on the importance of a real-time resource tracking for improving construction performance. To address this need, different tracking technologies such as barcodes and RFID tags [14–19], Ultra WideBand (UWB) [20–22,61–63,65,68], 3D range imaging cameras [21,23], global and local positioning systems (GPS) [21,23,64], and computer vision techniques [24,25,66,67,69–75] have been tested to provide tracking data for onsite construction resources. While dominantly used for tracking construction material, they have also been used in locating workers and recording the sequence of their movement necessary to complete a task; e.g., [2,6,24–28,59,60,72–76]. For the task of performance monitoring, there is a need for detailed data on activities of construction equipment and workers, which makes a low-cost vision-based method, an alternative appealing solution; particularly because a low-cost single camera (e.g., $40–100 Wi-Fi HD camera) can potentially be used for (1) recognizing activities of multiple equipment and workers for both performance monitoring and safety analysis purposes and (2) minimizing the need for sophisticated on-board telemetric sensory for each equipment (or other sensory mentioned above for each worker) which can come at a higher cost.

2.1. Construction equipment action recognition

Despite a large number of emerging works in the area of human *action recognition* for smart online queries or robotic purposes and their significance for performance assessment on construction sites, this area has not yet been thoroughly explored in the Architecture/Engineering/Construction (AEC) community. The work in [13] is one of the first in this area, which presented a vision-based tracking model for monitoring a tower crane bucket in concrete placement operations. Their proposed method is focused on action recognition of crane buckets and hence it cannot be directly applied to earthmoving operations. In a more recent work, Gong and Caldas [2] proposed an action recognition method based on an unsupervised learning algorithm and showed promising results. However, generalizing the applicability of unsupervised learning models for unstructured construction sites can be challenging. In this paper we show that a supervised learning method may provide better performance in the equipment action recognition task. Zou and Kim [6] also presented an image-processing approach that automatically quantifies the idle time of a hydraulic excavator. The approach uses color information for detecting motion of equipment in 2D and thus can be challenged by changes of scene