Automation in Construction 24 (2012) 194–202

Contents lists available at SciVerse ScienceDirect

Automation in Construction

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

Part based model and spatial–temporal reasoning to recognize hydraulic excavators in construction images and videos

Ehsan Rezazadeh Azar *, Brenda McCabe

Department of Civil Engineering, University of Toronto, Toronto, ON, Canada, M5S 1A4

ARTICLE INFO

Article history:
Accepted 9 March 2012
Available online 5 April 2012

Keywords:
Computer vision
Object recognition
Earthmoving equipment
Automated data collection
Digital images
Spatial–temporal reasoning

ABSTRACT

Detection of earthmoving equipment in construction images and videos can increase the automation level of many construction management tasks such as productivity measurement, locating of machines, work-zone safety, and semantic image and video indexing. Some of the earthmoving plants, such as hydraulic excavator, have articulated shapes making them a difficult target for even state of the art object recognition algorithms. The goal of this paper is to develop a model for non-rigid equipment detection and pose estimation in construction images and videos. In this paper, we describe an object recognition system based on mixture of appearances of deformable body parts of the hydraulic excavator and compare its results with general Histogram of Oriented Gradient detectors in both images and videos. Then a spatial–temporal reasoning model is presented which uses time and space constraints of the excavators' moving patterns to improve the detection results in videos.

1. Introduction

A variety of earthmoving equipment are used in heavy civil construction. Many of these plants are manufactured specifically to carry out one type of operation, while others, such as hydraulic excavators, can handle multiple activities, in this case excavation, loading, trimming, and moving materials. To improve the productivity of these costly resources, contractors generally try to minimize their idle time and non-value adding activities [1]. While smaller projects can be effectively managed using experienced site managers, larger projects benefit from the use of simulation or other methods to optimize resource utilization and therefore productivity. However, planning methods such as simulation, linear programming, and genetic algorithms do not always accurately predict the actual production rates. Human factors, equipment breakdowns, continually changing travel routes, degradation of the quality of site roads, and weather conditions are a few of the factors affecting the earthmoving operations. As a result, there has been growing interest in methods to monitor the operations in real-time to find discrepancies between the planned and real performance, record productivity data for future projects, and to make timely changes to improve the operations as quickly as possible [2]. Automated vehicle tracking technologies such as global positioning system (GPS) [3,4] and ultra wideband (UWB) [5] have been used to automate data collection. These real-time positioning devices provide the three-dimensional location of the equipment, which is useful to interpret the activities of mobile machines such as dump trucks and rollers [3,4]; even though they cannot effectively distinguish productive activities from non value-added (NVA) traverses. These tools can also locate the stationary plants such as excavators, which provide a valuable data for planning and management of these resources; however, they cannot distinguish whether the equipment is idle or active. Various built-in sensing devices can provide a wide range of data from the machine itself such as engine operating parameters [6], body, boom, and bucket orientations [7] to facilitate operating the equipment efficiently. It is also possible to collect and interpret these data to measure the productivity of the machine. However, there are some limitations with these devices related to cost-effectiveness and data interpretation. In addition, all of the mentioned devices have issues in rented equipment, which is commonly employed by general contractors, because of the effort and cost to continually install and remove sensing tags from the equipment and update the monitoring software. Leaving technology aside, manual optimization through continuous supervision of the operation is labor-intensive, expensive, and error prone.

Computer-assisted visual monitoring has a potential to detect, track, and measure the productivity of stationary and mobile equipment. Because major projects generally take place in open field and mining sites, sight lines that are not obscured can be selected. Although many construction sites already have surveillance cameras that capture videos or time lapsed photos at regular time intervals [1,8], they are not often used to their full potential due to the labor-intensive process of manually extracting data from the images and videos before they are deleted in a rotation to save memory. Computer vision algorithms are now being evaluated to automate data extraction from construction

* Corresponding author. Tel.: +1 416 8317631; fax: +1 416 978 5054.
E-mail addresses: rezazade@ecf.utoronto.ca (E. Rezazadeh Azar), brenda.mccabe@utoronto.ca (B. McCabe).

0926-5805/$ – see front matter
doi:10.1016/j.autcon.2012.03.003

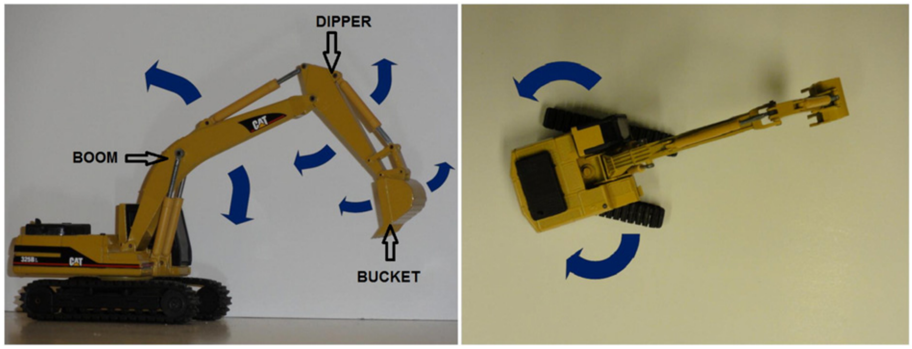

Fig. 1. Deformations of the hydraulic excavators.

videos for tracking resources and personnel [9,10], measuring concrete pouring cycles [8], and assessing human workforce productivity [11,12].

This paper describes part of an extensive research effort to develop an automated vision-based system to monitor earthmoving activities. The envisioned system would have three main functions: object detection, object tracking, and activity recognition. The goal of object recognition is to find and identify the equipment on site from video frames. Existing object detection algorithms such as Histogram of Oriented Gradients (HOG) [13] show promising performance in recognition of rigid objects such as urban vehicles [14] and off-highway dump trucks as an example of rigid construction equipment [15]; however, deformable object detection and articulated pose estimation are more challenging problems faced by the computer vision community [16,17]. A number of earthmoving equipment have non-rigid frames that move or deform their parts to perform the work, including hydraulic excavators, loaders, and articulated dump trucks.

Hydraulic excavators are highly articulated and are extensively used in construction due to their versatility to carry out a variety of operations in different environments from quarries to compact urban sites. They have also been the subject of other research such as automating excavator operation [18], and improving equipment safety [19]. Measuring the idle time of hydraulic excavators using color space values [1] to isolate the machine in images with soil and snow backgrounds was attempted. In another research, moving entities in construction videos were segmented by a background subtraction filter, and the remaining blobs were classified with respect to four features including aspect ratio, height-normalized area size (occupied area size in pixels/centroid of the area height), percentage of occupancy of the bounding box, and average gray-scaled color of the area [20].

In this paper we firstly describe the cutting edge methods currently used to detect deformable objects, and then present our part-based recognition algorithm, which was inspired by those methods. Afterwards, we employ spatial–temporal constraints of the moving patterns of the excavators to eliminate false detections and validate true positives. Finally, the evaluation results of the algorithms on a test dataset are presented.

2. Articulated object recognition

Object recognition in digital images is challenging due to changing viewpoints, illumination, occlusions, and scale. Robust algorithms have been developed in the last decade to detect rigid objects such as Histogram of Oriented Gradients (HOG) [13] and Haar-like features [21]. In HOG algorithm, which was one of the most successful rigid object detection methods, computed gradients of the gray-scaled image are discretized into spatial and orientation cells to form a descriptor vector. Then positive and negative vectors from a training dataset are trained by the SVMLight method [22] resulting in a single classifying vector. This detector uses a sliding window technique to check all locations and scales of an image. If the result exceeds a determined threshold, that window is accepted as a target.

Dramatically more difficult to detect are objects that change shape, and current research is focused on human detection and pose identification due to the complicated configurations of the human body. The outcomes are highly applicable for security surveillance, traffic safety, and image indexing. Many of the recently developed models use part-based and pictorial structures to find a group of parts of a semantic object arranged in a deformable configuration [16,17]. One of the cutting edge algorithms is the latent support vector machine (Latent SVM) part based models [17], which won the 2009 PASCAL object detection challenge [23]. This model uses a modified HOG detector, called a root filter, to locate the most probable candidates for the object within the image. It then searches for the parts of the object at twice the spatial resolution relative to the features captured by the root filter inside the detected root areas. In the current research, we used a similar idea with substantial modifications to detect a root and then search for the possible configurations of the parts of the excavator to both detect and estimate the pose of the equipment.

3. Research methods

3.1. Deformable parts

Hydraulic excavators are highly deformable machines that can slew 360° and rotate all three parts of their arm (boom, dipper, and attachment) around their hinged supports. The typical deformations of the machine are illustrated in Fig. 1. As a result, the machine can have countless forms, making it impossible to be detected with a limited number of training configurations as used in the case of rigid-frame equipment [14,15].

In latent SVM part-based models [17], the first classifier is trained to detect the entire body (e.g. human), which is called the root. Then it searches for the body parts (e.g. torso, arms, and legs) inside the

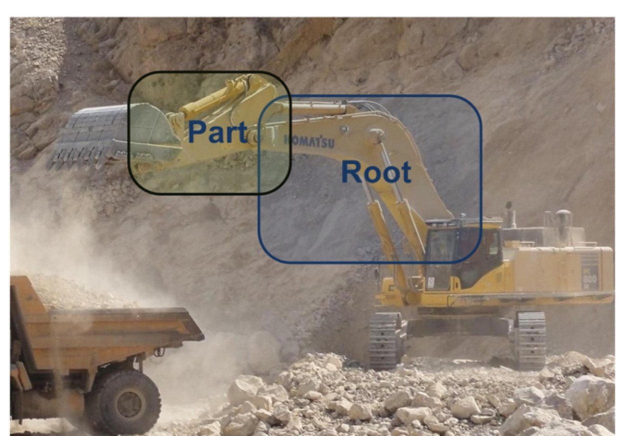

Fig. 2. Root and part of the excavator.