Automation in Construction 96 (2018) 148–160

Contents lists available at ScienceDirect

Automation in Construction

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

Camera marker networks for articulated machine pose estimation

Chen Feng[a,*], Vineet R. Kamat[a], Hubo Cai[b]

[a] Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, USA
[b] School of Civil Engineering, Purdue University, West Lafayette, IN, USA

ARTICLE INFO

Keywords:
Camera
Marker
Excavator pose estimation
Error analysis
Structure-from-motion

ABSTRACT

The pose of an articulated machine includes the position and orientation of not only the machine base (e.g., tracks or wheels), but also its major articulated components (e.g., stick and bucket). To automatically estimate this pose is a crucial component of technical innovations aimed at improving both safety and productivity in many construction tasks. Based on computer vision, an automatic observation and analysis platform using a network of cameras and markers is designed to enable such a capability for articulated machines. To model such a complex system, a theoretical framework termed camera marker network is proposed. A graph abstraction of such a network is developed to both systematically manage observations and constraints, and efficiently find the optimal solution. An uncertainty analysis without time-consuming simulation enables optimization of network configurations to reduce estimation uncertainty, leading to several empirical rules for better camera calibration and pose estimation. Through extensive uncertainty analyses and field experiments, this approach is shown to achieve centimeter level bucket depth tracking accuracy from as far as 15 m away with only two ordinary cameras (1.1 megapixels each) and a few markers, providing a flexible and cost-efficient alternative to other commercial products that use infrastructure dependent sensors like GPS. A working prototype has been tested on several active construction sites confirming the method's effectiveness.

1. Introduction

The construction industry has long been affected by high rates of workplace injuries and fatalities. According to the United States Bureau of Labor Statistics' 2013 Census of Fatal Occupational Injuries (CFOI) report [1], the construction industry had the largest number of fatal occupational injuries, and in terms of rate ranked the fourth highest among all industries.

In addition to the safety concerns, there are also increasing concerns of relatively stagnant productivity rates and skilled labor shortage in the construction industry. For example, recently the construction sector in the United Kingdom is reported to be in urgent need of 20% more skilled workers and thus 50% more training provision by 2017, to deliver projects in planning [27].

Excavation is a typical construction activity affected by the safety and productivity concerns mentioned above. Excavator operators face two major challenges during excavation operations, described as follows.

First is *how to maintain precise grade control*. Currently, grade control is provided by employing grade-checkers to accompany excavators during appropriate operations. Grade-checkers rely on surveying and frequently monitor the evolving grade profile. The evolving grade profile is compared to the target grade profile and this information is communicated by the grade-checker to the excavator operator. The operator reconciles this information and adjusts the digging strokes accordingly. This process is repeated until the target profiles are achieved. Employing grade-checkers is not only dangerous but also results in a significant loss in excavation productivity due to frequent interruptions required for surveying the evolving profile.

Second is *how to avoid collisions* to either human workers, buried utilities, or other facilities, especially when excavator operators cannot perceive the digging machine's position relative to hidden obstructions (i.e., workers or utilities) that it must avoid. According to the aforementioned CFOI report, among all the causes for the 796 fatal injuries in the U.S. construction industry in 2013, the cause of striking by object or equipment comprised 10%. This percentage is even higher in other industries such as agriculture (19%), forestry (63%), and mining (23%). Besides directly causing fatal injuries on jobsites, construction machines can also inadvertently strike buried utilities, thus disrupting life and commerce, and pose physical danger to workers, bystanders, and building occupants. Such underground strikes happen with an average frequency of about once per minute in the U.S., reported by the

---

[*] Corresponding author.
E-mail address: cforrest@umich.edu (C. Feng).

https://doi.org/10.1016/j.autcon.2018.09.004
Received 4 May 2017; Received in revised form 28 July 2018; Accepted 5 September 2018

Common Ground Alliance,[1] the nation's leading organization focused on excavation safety. More specifically, excavation damage is the third biggest cause of breakdowns in U.S. pipeline systems, accounting for about 17% of all incidents, leading to over 25 million annual utility interruptions [34].

Automation and robotics in construction (ARC) has been extensively promoted in the literature as a means of improving construction safety, productivity and mitigating skilled labor shortage, since it has the potential to relieve human workers from either repetitive or dangerous tasks and enable a safer collaboration and cooperation between construction machines and the surrounding human workers. In order to apply ARC and increase intelligence of construction machines to improve either safety or productivity for excavation and many other activities on construction jobsites, one of the fundamental requirements is the ability to automatically and accurately estimate the pose of an articulated machine (e.g., excavator or backhoe). The pose here includes the position and orientation of not only the machine base (e.g., tracks or wheels), but also each of its major articulated components (e.g., stick and bucket).

When a construction machine can continuously track its end-effector's pose on the jobsites, such information can be combined together with the digital design of a task, either to assist human operators to complete the task faster and more efficiently, or to eventually finish the task autonomously. For example, an intelligent excavator being able to track the pose of its bucket can guide its operator to dig trenches or backfill according to designed profiles more easily and accurately with automatic grade-check. This can eventually lead to fully autonomous construction machines. When construction machine becomes more intelligent, it can be expected to save time in training operators and thus to mitigate skilled labor shortage and also improve productivity.

On the other hand, when construction machines are aware of the poses of their components at any time and location on jobsites, combined with other abilities such as the recognition of human workers' poses and actions, such machines will be able to make decisions to avoid striking human workers, for example by sending alerts to their operators or even temporarily taking over the controls to prevent accidents. Thus, it will help to decrease the possibilities of those injuries and fatalities and improve the safety on construction jobsites. Similarly, with continuous tracking of the pose of its end-effector (e.g., a bucket of an excavator), an intelligent excavator could perform collision detection with an existing map of underground utilities and issue its operator a warning if the end-effector's distance to any buried utilities exceeds some predefined threshold.

Thus, from a safety, productivity, and economic perspective, it is critical for such construction machines to be able to automatically and accurately estimate poses of any of their articulated components of interest. In this paper, a computer vision based solution using planar markers is proposed to enable such capability for a broad set of articulated machines that currently exist, but cannot track their own pose. A working prototype is implemented and shown to enable centimeter level excavator bucket depth tracking. Its overview is shown in Fig. 1, with (A) camera cluster and stick marker, (B) benchmark with pre-surveyed pose in the project reference frame, (C) system calibration, (D) working prototype of automatic grade control, (E) comparison to manual grade.

The remainder of the paper is organized as follows: Related work is reviewed in Section 2. A detailed introduction and analysis of the proposed camera marker networks are explained in Section 3. The articulated machine pose estimation system based on the theoretical framework is discussed in Section 4. The experimental results are presented and discussed in Section 5. Finally, the conclusions are drawn and the authors' future work is summarized in Section 6.

2. Previous work

The majority of the construction machines on the market do not have the ability to track their poses relative to some project coordinate frames of interest. To track and estimate the pose of an articulated machine, there are mainly four groups of methods.

First are the 2D video analysis methods, stimulated by the improvement in computer vision on object recognition and tracking. Static surveillance cameras were used to track the motion of a tower crane in Yang et al. [2] for activity understanding. Similarly in Rezazadeh Azar and McCabe [3] part based model was used to recognize excavators for productivity analysis. This type of methods generally requires no retrofitting on the machine, but suffers from both possibilities of false or missed detection due to complex visual appearance on jobsites and the relative slow processing speed. Although real-time methods exist as in [18,30], they either cannot provide accurate 6D pose estimation, or require additional information such as a detailed 3D model of the machine.

Second are stereo vision based methods. A detailed 3D model of the articulated object was required in Hel-Or and Werman [4] in addition to stereo vision. A stereo camera was installed on the boom of a mining shovel to estimate pose of haul trucks in Borthwick et al. [17], yet the shovel's own pose was unknown. In Lin et al. [5] the shovel's swing rotation was recovered using stereo vision SLAM, yet the pose of its buckets was not estimated. This type of methods can be infrastructure independent if with SLAM, yet some problems (sensitivity to lighting changes or texture-less regions) remain to be resolved for more robust applications.

Third are laser based methods, e.g., [20,22,25] which rooted from the extensive use of laser point clouds in robotics. This type of methods can yield good pose estimation accuracy if highly accurate dense 3D point clouds of the machine are observed using expensive and heavy laser scanners. Otherwise with low quality 2D scanners, only decimeter level accuracy was achieved [25].

Finally are angular sensor based methods, such as [6–8]. They are usually infrastructure independent and light-weight, but the resulting pose estimation is either not accurate enough or sensitive to changes of magnetic environment which is not uncommon in construction sites and can lead to large variations in the final estimation of the object poses. Moreover, this type of methods only estimate the articulated machine's pose relative to the machine base itself, if without the help of infrastructure dependent sensors like GPS. However, the use of GPS brings several technical challenges. For example, construction sites in a lot of cases do not have good GPS signals to provide accurate position estimation when these sites are located in urban regions or occluded by other civil infrastructure such as under bridges. Sometimes GPS signals could even be blocked by surrounding buildings on jobsites and thus fail to provide any position estimation. In addition, since the GPS only provides 3D position estimation, to get the 3D orientation estimation one needs at least two GPS receivers at different locations of a rigid object. When the object is small, such as a mini-excavator's bucket, the estimated 3D orientation's uncertainty will be high.

Our method is different than the above groups of methods in several aspects. Firstly, we use 2D cameras, instead of stereo cameras or other non-vision sensors, as our input sensors. Secondly, our method aims to estimate accurate 6D poses instead of simply tracking 2D positions on an image, thus providing more valuable information for downstream applications such as control or collision avoidance. Thirdly, we introduce fiducial markers to augment the working environment so as to increase the robustness of the system, whereas other non-marker-based vision methods might suffer from challenges such as texture-less environment. This is based on our argument that trade-off should be make between deployment efforts versus system robustness and accuracy. This paper significantly extends our introductory ISARC 2015 paper [9] describing the system, and discusses the development of a generalized theory for machine pose estimation along with an extended set of

---

[1] http://www.commongroundalliance.com.