Vision-Based Articulated Machine Pose Estimation for Excavation Monitoring and Guidance

C. Feng, S. Dong, K.M. Lundeen, Y. Xiao, and V.R. Kamat

Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, USA
E-mail: cforrest@umich.edu, dsuyang@umich.edu, klundeen@umich.edu, yongxiao@umich.edu, vkamat@umich.edu

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 each of its major articulated components (e.g., stick and bucket). The ability to automatically estimate this pose is a crucial component of technical innovations aimed at improving both safety and productivity in many construction tasks. A computer vision based solution using a network of cameras and markers is proposed in this research to enable such a capability for articulated machines. Firstly, a planar marker is magnetically mounted on the end-effector of interest. Another marker is fixed on the jobsite whose 3D pose is pre-surveyed in a project coordinate frame. Then a cluster of at least two cameras respectively observing and tracking the two markers simultaneously forms a camera-marker network and transfers the end-effector's pose into the desired project frame, based on a pre-calibration of the relative poses between each pair of cameras. Through extensive sets of uncertainty analyses and field experiments, this approach is shown to be able to achieve centimeter level depth tracking accuracy within up to 15 meters with only two ordinary cameras (1.1 megapixel 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 with positive feedback from excavator operators confirming the solution's effectiveness.

Keywords -

Pose Estimation, Camera-Marker Network, Bundle Adjustment, Uncertainty Analysis, Excavation Guidance

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.

Figure 1. SmartDig: (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

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

___
[1] http://www.bls.gov/iif/oshcfoi1.htm

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[2].

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 specialize in 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 percent. 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 Common Ground Alliance[3], 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[4].

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 machines 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 (Figure 1) is implemented and shown to enable centimeter level excavator bucket depth tracking.

The remainder of the paper is organized as follows: Related work is reviewed in section 2. The authors'

[2] http://www.kpmg.com/uk/en/issuesandinsights/articlespublications/pages/construction-skills-index-2014.aspx
[3] http://www.commongroundalliance.com/
[4] http://primis.phmsa.dot.gov/comm/reports/safety/CPI.html

technical approach is discussed next in detail in section 3. The experimental results are presented in section 4. Finally, in section 5, the conclusions are drawn and the authors' future work is summarized.

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 [1] for activity understanding. Similarly in [2] part based model was used to recognize excavators for productivity analysis. This type of methods generally require 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 [3, 4], 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 [5] in addition to stereo vision. A stereo camera was installed on the boom of a mining shovel to estimate pose of haul trucks in [6], yet the shovel's own pose was unknown. In [7] 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., [8, 9, 10], 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 [9].

Finally are angular sensor based methods, such as [11, 12, 13]. 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 sensors dependent on infrastructure that consume power and need careful maintenance 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 [14]. 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.

3   Technical Approach

In this section, different versions of the proposed articulated machine pose estimation system design are explained first. Then, the process to calibrate this system is described. Finally, uncertainty analysis is explored for the system with some important observations of the relationship between the system configuration and its stability, i.e., uncertainty of the estimated pose.

3.1   System Design

As mentioned previously, this computer vision based articulated machine pose estimation solution relies on a method called marker based pose estimation. Generally, marker based pose estimation firstly finds a set of 2D geometry features (e.g., points or lines) on an image captured by a calibrated camera, then establishes correspondences between another set of 2D or 3D geometry features on a marker whose pose is known with respect to a certain coordinate frame of interest, and finally estimates the pose of the camera in that coordinate system. If 2D-2D correspondences are used, the pose is typically estimated by homography decomposition. If 2D-3D, the pose is typically estimated by solving the perspective-n-point (PnP) problem. Two typical marker-based pose estimation methods are AprilTag [15] and KEG [16] algorithms.

There are two ways of applying marker based pose estimation for poses of general objects of interest. As shown in Figure 2, one way is to install the calibrated camera 1 rigidly on the object of interest (in this case, the cabin of the excavator), and pre-survey the marker 1's pose in the project coordinate frame. The other way is to install the marker 2 rigidly on the object (in this case, the stick of the excavator), and pre-calibrate the camera 2's pose in the project coordinate frame. As long as the camera 2 (or the marker 1) stays static in the project coordinate frame, the pose of the excavator's stick (or the cabin) can be estimated in real-time.
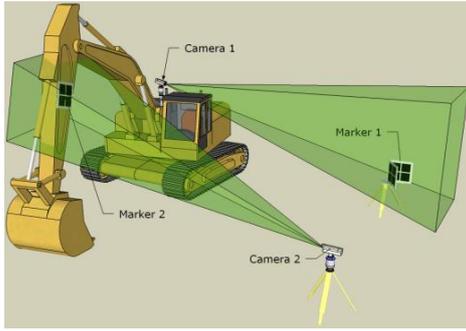
Figure 2. Two examples of basic camera-marker configuration

However, these basic configurations don't always satisfy application requirements. For example, if only the camera 1 and the marker 1 are used, the excavator's stick pose cannot be estimated. On the other hand when only the camera 2 and the marker 2 are used, once the stick leaves the field of view (FOV) of the camera 2, the stick's pose becomes unavailable as well. Thus it is necessary to take a camera's FOV into consideration when designing an articulated machine pose estimation system. This understanding leads to the camera-marker network design proposed as follows:

3.1.1 Camera-Marker Network

A camera-marker network is an observation system containing multiple cameras or markers for estimating poses of objects embedded in this system. It can be abstracted as a graph with three types of nodes and two types of edges (e.g., Figure 3). A node denotes an object pose (i.e. the local coordinate frame of that object), which can be a camera, a marker, or the world coordinate frame. An edge denotes the relative relationship between two objects connected by this edge, which can be either image point observations for the previously mentioned marker based pose estimation, or a known pose constraint (e.g., through calibration).
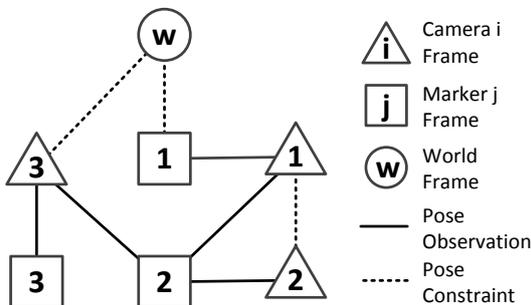
Figure 3. An example graph of a camera-marker network

Thus, if at least one path exists between any two nodes in such a graph, the relative pose between them can be estimated. In addition, any loop in the graph means a constraint of poses that can be used to improve the pose estimation. For example, in Figure 3, marker 2's pose in the world frame can be found via a path through camera 3 whose own pose in the world frame is pre-calibrated. The marker 2's pose can also be better estimated when observed by the rigidly connected camera 1 and 2 whose relative pose is pre-calibrated, since a loop is created.

Applying this concept to articulated machine pose estimation results in numerous possible designs. One of the possible camera-marker networks is shown in Figure 4, camera 1 observes the benchmark while camera 2 observes the stick marker, and the rigid transformation between the two cameras is pre-calibrated. Thus as long as the two markers stay inside the two cameras' FOV respectively, the stick's pose in the world frame can be estimated. It is worth noting that this only illustrate a simple configuration. With more cameras and markers in the network, there are more chances of creating loops and thus improving pose estimation, especially considering that surveillance cameras are becoming popular in construction jobsites whose poses can be pre-calibrated and thus act as the camera 3 in Figure 3.
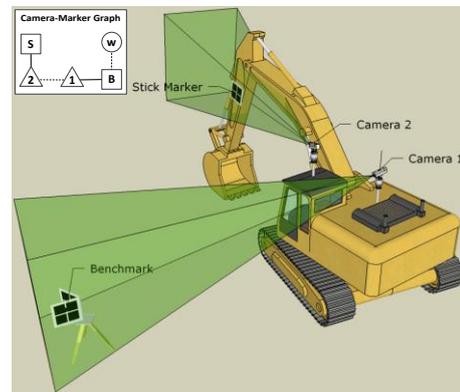
Figure 4. Multiple-camera multiple-marker configuration

3.1.2 Prototypes

Multiple prototypes have been implemented to realize the above described camera-marker network designs. Figure 5 demonstrates one of the early prototypes implementing a single-camera multiple-marker configuration. A mechanical component using a timing belt was adopted to map the relative rotation between the excavator bucket and the stick to the relative rotation between the stick marker and the flip marker. This implementation enables pose tracking of the excavator bucket.
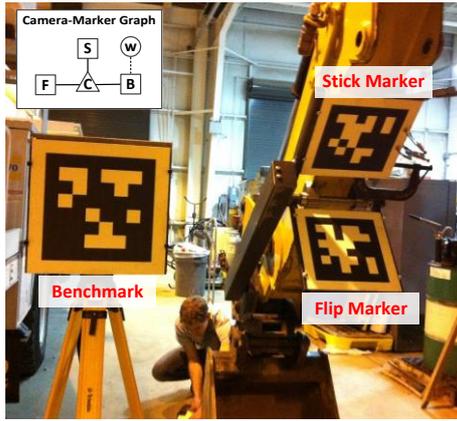
Figure 5. An early prototype configuration

Due to the potential interference of the flip marker and any obstructions during excavation, the above early prototype was slightly modified and evolved to the current prototype as shown in Figure 1. The newer working prototype implements the multiple-camera multiple-marker configuration similar to Figure 4. Two cameras are rigidly mounted forming a camera cluster. A linear potentiometer is installed on the stick to track the relative motion of the excavator bucket and the stick even if the bucket is deep inside the earth.

3.2 System Calibration

Two types of calibration are necessary for an articulated machine pose estimation system implementing the above camera-marker network design.

The first type is intrinsic calibration which determines internal parameters (e.g., focal length) of all cameras in the system. This is done using same methods as in [17].

The second type is extrinsic calibration which determines relative poses (e.g. dotted edges in the graph) designed to be calibrated before system operation. There are two kinds of such poses: 1) poses of static markers in the world frame, and 2) poses between rigidly connected cameras. The first kind of poses can be calibrated by traditional surveying methods using a total station. The second kind of poses, however, cannot be directly surveyed physically since a camera frame's origin and principal directions usually cannot be found or marked tangibly on that camera. Thus to calibrate a set of m rigidly connected cameras, a camera-marker graph needs to be constructed as denoted in Figure 6. A set of n markers' poses need to be surveyed in the world frame. Then when the m cameras observe these n calibration markers, the graph is formed to estimate each camera's pose in the world frame and thus their relative poses between each other (i.e., edges with question mark) are calibrated. It is suggested to ensure

that multiple loops exist in this graph to improve the accuracy of the poses to be calibrated. Such loop exists as long as at least two markers are observed by a same camera simultaneously. It is also worth noting that with enough many calibration markers, each camera's intrinsic parameters can be further optimized together with their extrinsic parameters.
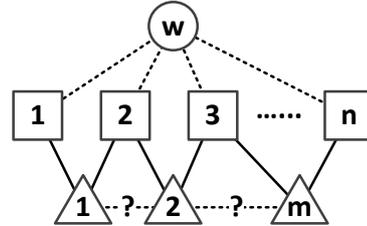
Figure 6. A camera-marker graph for extrinsic calibration

3.3 Uncertainty Analysis

It is not sufficient to only estimate the pose of an articulated machine. The uncertainty of the estimated pose is critical for the following reasons. Firstly the uncertainty provides a measure of the confidence level of the estimated pose, which is necessary for many downstream applications (e.g., deciding buffer size for collision avoidance). Secondly it serves as a tool for evaluating the stability of the pose estimation system under different configurations, and thus further guiding to avoid critical configurations that lead to unstable pose estimation.

To perform uncertainty analysis on the proposed camera-marker network pose estimation system, the system is firstly abstracted as the following state space model:

$$\mathbf{Z} = F(\mathbf{X}; \mathbf{Y}, \mathbf{C}) \qquad (1)$$

where $\mathbf{X}$ is the state vector of the network (usually encodes the poses of nodes in the graph), $\mathbf{Z}$ is the predicted measurement vector containing image coordinates of all the points projected from markers, $\mathbf{Y}$ is the known parameters (usually contains marker points' local coordinates), $\mathbf{C}$ is the calibrated parameters (usually encodes all cameras' intrinsic parameters and all calibrated poses), and F is the system's observation function parameterized by $\mathbf{Y}$ and $\mathbf{C}$, i.e., the camera perspective projection function.

For example, for a network of a single camera and a single marker, $\mathbf{X}$ is a 6×1 vector that encodes the marker's orientation and position in the camera frame; $\mathbf{Y}$ is a $3n \times 1$ vector containing n marker points' coordinates from surveying; $\mathbf{C}$ is a vector of the camera intrinsic parameters. If another marker is added to this

network, **Y** should be extended with points on the new marker.

3.3.1 Uncertainty Propagation

No matter how complex such a network is and what method is used to get an initial estimate of **X** (either PnP or homograph decomposition), the optimized state $\hat{\mathbf{X}}$ can be calculated by the following least square optimization, i.e., bundle adjustment:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \left\| \hat{\mathbf{Z}} - F(\mathbf{X}; \mathbf{Y}, \mathbf{C}) \right\|_{\mathbf{P}_{\hat{\mathbf{Z}}}}^2 \qquad (2)$$

where $\mathbf{P}_{\hat{\mathbf{Z}}}$ is the a priori covariance matrix of the actual measurements $\hat{\mathbf{Z}}$, typically assumed as $\sigma_u^2 \mathbf{I}$ when image coordinates are measured with a standard deviation of σ_u.

To backward propagate the measurement uncertainty $\mathbf{P}_{\hat{\mathbf{Z}}}$ to the uncertainty of the optimized state $\hat{\mathbf{X}}$ requires linearization of F around $\hat{\mathbf{X}}$. Since the error is assumed to come from only the measurements (the uncertainty in calibrated parameters **C** can be included in future work, but is assumed to be negligible in this paper), one can directly apply the results in [18] to calculate the uncertainty of the optimized states:

$$\mathbf{P}_{\hat{\mathbf{X}}} = (\mathbf{J}^{\mathbf{T}} \mathbf{P}_{\hat{\mathbf{Z}}}^{-1} \mathbf{J})^{-1} = \sigma_u^2 (\mathbf{J}^{\mathbf{T}} \mathbf{J})^{-1} \qquad (3)$$

where $\mathbf{J} = \dfrac{\partial F}{\partial \mathbf{X}} \Big|_{\hat{\mathbf{X}}}$ is the Jacobian matrix of F evaluated at $\hat{\mathbf{X}}$.

3.3.2 Uncertainty and Configuration

Equation (3) not only provides a means of evaluating uncertainty of the optimized pose estimation of a camera-marker network, but also provides a tool to predict the system stability at any given system configuration **X** before even making any measurements. This is done by evaluating the Jacobian matrix **J** of F at that **X**, and then applying equation (3) to predict the covariance matrix. It is based on the fact that the aforementioned backward propagation of measurement uncertainty does not directly rely on specific measurements. In fact it directly relies on the system configuration **X** around which the linearization is performed. Thus, when evaluating Jacobian matrix **J** at a configuration **X**, equation (3) yields the theoretically best/smallest pose estimation uncertainty one can expect at that configuration, which denotes the system stability at that configuration.

Using this method, some important empirical

conclusions on the basic *single-camera single-marker system* are found about relationships between system stability and configuration, based on numerical experiments, which are useful for more complex system design and are listed as follows. Similar analysis will be performed to multiple-camera or multiple-marker system in future work.

1. *The marker's origin/position in the camera frame, $^c\mathbf{t}_m$, has the largest uncertainty along a direction nearly parallel to the camera's line of sight to the marker*, i.e., $^c\mathbf{t}_m$ itself. Figure 7 exemplifies this observation at two randomly generated poses between the camera and the marker.
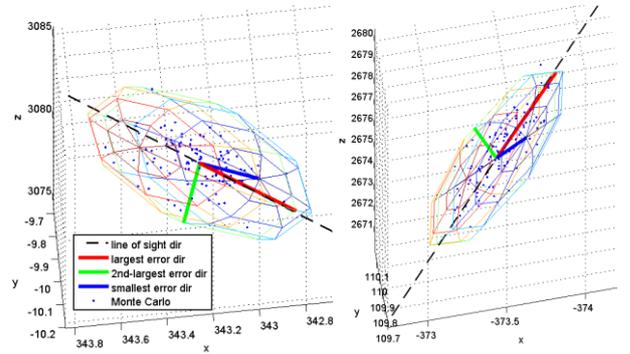
Figure 7. Largest position error direction

2. *The largest uncertainty of marker's position in the camera frame increases approximately quadratic to the marker's distance to the camera; compared to which the two smallest uncertainty's increases are almost negligible.* Figure 8 shows a typical example.
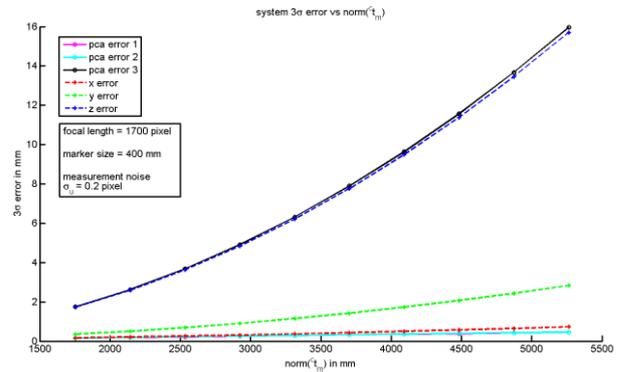
Figure 8. Position error vs marker distance

3. *The largest uncertainty of marker's position in the camera frame increases approximately linear to the camera focal length; compared to which the two smallest uncertainty's increases are almost negligible.* Figure 9 shows a typical example.
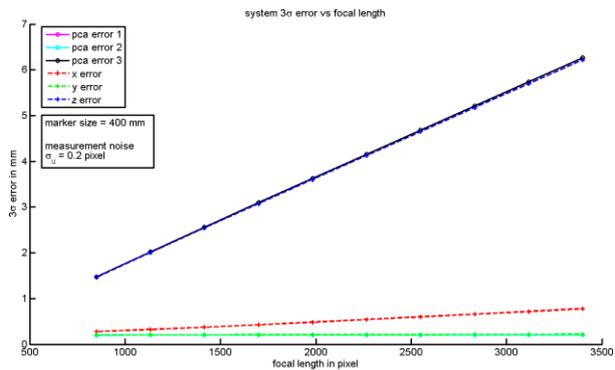
Figure 9. Position error vs focal length

4 Experimental Results

4.1 Feasibility Experiments

Before implementing the pose estimation system prototypes, a set of experiments were performed to test the feasibility of marker based pose estimation in different indoor/outdoor construction environments. In all the experiments, AprilTag [15] was chosen as the basic marker detection and tracking algorithm.

Firstly, the outdoor detectability of markers was tested. A marker's detectability is a function of many factors including the marker size, the distance between the marker and the camera, included angle between the camera viewing direction and the marker plane's normal direction, and also image resolution. Since the distance between the marker and the camera is the most critical factor affecting the method's feasibility in real applications, this experiment is performed by fixing other factors and then gradually increasing the distance of the marker in front of the camera, until the algorithm fails to detect the marker, and recording the distance. Varying other factors and repeating this process results in Table 1. One can consult this table to decide how large the marker should be to fit application need.

Secondly, illumination is a critical factor affecting performance of many computer vision algorithms. The AprilTag algorithm was thus tested under various illumination conditions to examine its robustness for construction applications. Figure 10 shows successful marker detection under different indoor/outdoor lighting conditions. These experiments and following extensive prototype tests proved AprilTag based marker detection method's robustness to illumination changes.
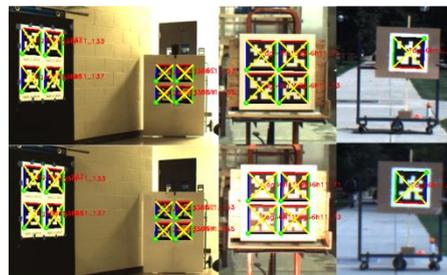
Figure 10. Marker detection vs illumination

Finally, for uncertainty propagation, one needs to have a prior estimation of the image measurement noise's standard deviation σ_u. This is achieved by collecting multiple images under a static camera marker pose. Repeating this process for different poses and collecting corresponding image measurement statistics lead to an image measurement covariance matrix Σ_u, which can be further relaxed to $\sigma_u^2 \mathbf{I}$ to include all the data points. Figure 11 shows that $\sigma_u = 0.2$ pixel is reasonable.
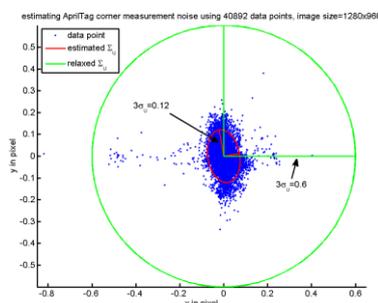
Figure 11. Image measurement noise estimation

Table 1. Outdoor detectability of AprilTag

Max Detectable Distance (m)		Marker Angle (degree)			
		0	45	0	45
Marker Size (m²)	0.2 x 0.2	6.10	4.88	11.28	8.84
	0.3 x 0.3	8.23	7.01	14.94	11.58
	0.46 x 0.46	13.41	11.28	25.91	21.64
	0.6 x 0.6	19.51	16.46	34.44	30.48
Image Resolution		640 x 480		1280 x 960	
Focal Length		850 pixels		1731 pixels	
Processing Rate		20 Hz		5 Hz	

4.2 Prototype Experiments

As previously mentioned, a multiple-camera multiple-marker articulated machine pose estimation prototype has been implemented with the application of estimating an excavator's bucket depth in a project frame, which could be used for automatic excavation guidance and grade control.

(a) Setup

(b) Different configurations
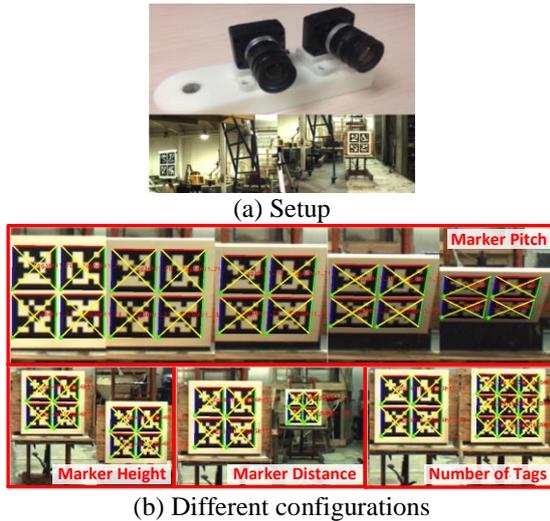
Figure 12. Prototype and experiment setup
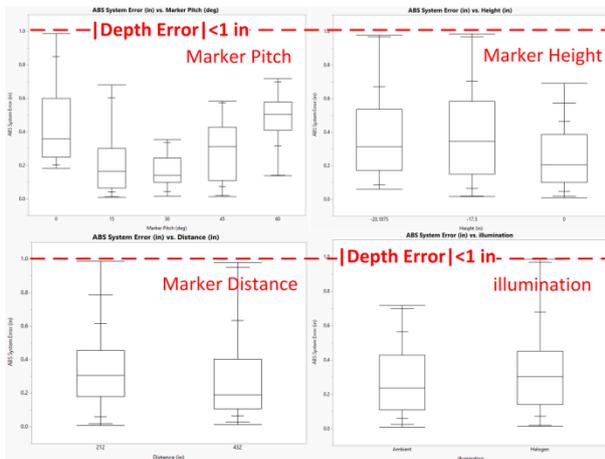
Figure 13. Prototype error vs. configuration

The top row of Figure 12(a) shows the camera cluster of the prototype in Figure 1, and different experiment configurations to test the depth estimate's accuracy. The experiments were setup by observing the two markers in the bottom row of Figure 12(a) using the two cameras in the cluster respectively. Then the depth difference between the two markers was estimated using the proposed method, while the ground truth depth difference between the two marker centers was measured by a total station with 1 mm accuracy. Figure 12(b) illustrates the configurations of different sets of such experiments, for comprehensive tests of the method's accuracy under several system and design variations. The first set varies one of the marker's pitch angle (top row of the figure). The second set varies its height (bottom-left). The third set varies its distance to the camera (bottom-middle). And the fourth set varies the number of tags used in that marker (bottom-right).

Figure 13 shows the absolute depth errors comparing the ground truths with the results from camera marker network pose estimation, in the above mentioned different sets of prototype experiments, using the box quartile plot. Note that **all errors are less than 2.54 cm, even when observed from more than 10 meters away**. Further experiments showed that the system worked up to 15 meters.

This prototype is then tested on a real excavation site for grade control as shown in Figure 1 (D) and (E). The resulting trench depth differences between the manual grade and the guided grade (by the prototype) are less than 1 inch, which fulfils the need of many construction applications.

5 Conclusion and Future Work

This paper proposed a vision based pose estimation solution for articulated machines using a camera-marker network. The uncertainty of the network pose estimation is analyzed through backward propagation of measurement covariance matrix. Based on this, an efficient approach of evaluating such a pose estimation system's uncertainty at any given state is introduced and applied to the basic single-camera single-marker system to find some important relationships between system states and corresponding system uncertainty, which is useful to guide more complex design. The conducted experiments and a working prototype proved the proposed solution's feasibility, robustness, and accuracy for real world construction applications.

The authors' current and planned work in this research direction is focused on continuously improving the estimation accuracy such as taking the uncertainty of calibrated parameters **C** into consideration, and also analyzing uncertainty versus system configuration for more complex camera-marker networks.

References

[1] J. Yang, P. Vela, J. Teizer and Z. Shi, "Vision-based crane tracking for understanding construction activity," *Proc. ASCE IWCCE,* pp. 258--265, 2011.

[2] E. Rezazadeh Azar and B. McCabe, "Part based model and spatial--temporal reasoning to recognize hydraulic excavators in construction images and videos," *Automation in construction,* vol. 24, pp. 194--202, 2012.

[3] M. Memarzadeh, A. Heydarian, M. Golparvar-Fard and J. Niebles, "Real-time and automated recognition and 2D tracking of Construction workers and equipment from Site video streams," in *Int. Workshop on Computing in Civil Engineering*, 2012.

[4] J. D. Brookshire, "Articulated pose estimation via over-parametrization and noise projection," MIT, Cambridge, 2014.

[5] Y. Hel-Or and M. Werman, "Model based pose estimation of articulated and constrained objects," in *ECCV*, 1994.

[6] J. R. Borthwick, P. D. Lawrence and R. H. Hall, "Mining haul truck localization using stereo vision," in *Proceedings of the Robotics and Applications Conference*, 2009.

[7] L.-H. Lin, P. D. Lawrence and R. Hall, "Robust outdoor stereo vision SLAM for heavy machine rotation sensing," *Machine vision and applications,* vol. 24, no. 1, pp. 205--226, 2013.

[8] E. Duff, "Tracking a vehicle from a rotating platform with a scanning range laser," in *Proceedings of the Australian Conference on Robotics and Automation*, 2006.

[9] A. H. Kashani, W. S. Owen, N. Himmelman, P. D. Lawrence and R. A. Hall, "Laser Scanner-based End-effector Tracking and Joint Variable Extraction for Heavy Machinery," *The International Journal of Robotics Research,* vol. 29, no. 10, pp. 1338-1352, 2010.

[10] Y. Cho and M. Gai, "Projection-Recognition-Projection Method for Automatic Object Recognition and Registration for Dynamic Heavy Equipment Operations," *Journal of Computing in Civil Engineering,* vol. 28, no. 5, p. A4014002, 2014.

[11] F. Ghassemi, S. Tafazoli, P. D. Lawrence and K. Hashtrudi-Zaad, "An accelerometer-based joint angle sensor for heavy-duty manipulators," in *Proceedings of IEEE International Conference on Robotics and Automation*, 2002.

[12] P. Cheng and B. Oelmann, "Joint-angle measurement using accelerometers and gyroscopes—A survey," *IEEE Transactions on Instrumentation and Measurement,* vol. 59, no. 2, pp. 404--414, 2010.

[13] S. Lee, M.-S. Kang, D.-S. Shin and C.-S. Han, "Estimation with applications to dynamic status of an excavator without renovation," *Gerontechnology,* vol. 11, no. 2, p. 414, 2012.

[14] Y. Cui and S. S. Ge, "Autonomous vehicle positioning with GPS in urban canyon environments," *IEEE Transactions on Robotics and Automation,* vol. 19, no. 1, pp. 15--25, 2003.

[15] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*, 2011.

[16] C. Feng and V. R. Kamat, "Plane Registration Leveraged by Global Constraints for Context-Aware AEC Applications," *Computer-Aided Civil and Infrastructure Engineering,* vol. 28, no. 5, pp. 325-343, 2012.

[17] C. Feng, Y. Xiao, A. Willette, W. McGee and V. R. Kamat, "Towards Autonomous Robotic In-Situ Assembly on Unstructured Construction Sites Using Monocular Vision," in *Proceedings of the 31th International Symposium on Automation and Robotics in Construction*, Sydney, Australia, 2014.

[18] R. Hartley and A. Zisserman, Multiple view geometry in computer vision, Cambridge University Press, 2000.

[19] C. Balaguer, "Nowadays trends in robotics and automation in construction industry: Transition from hard to soft robotics," in *Proceedings of International Symposium on Automation and Robotics in Construction*, Jeju. Korea, 2004.