A Point Cloud-Vision Hybrid Approach for 3D Location Tracking of Mobile Construction Assets

Y. Fang[a], J. Chen[b], Y. K. Cho[a], and P. Zhang[c]

[a] School of Civil and Environmental Engineering, Georgia Institute of Technology
[b] School of Electrical and Computer Engineering, Robotics Institute, Georgia Institute of Technology
[c] Department of Construction Management, Tsinghua University
E-mail: yihaifang@gatech.edu, jchen490@gatech.edu, yong.cho@ce.gatech.edu, zhangpy14@mails.tsinghua.edu.cn

Abstract –

Modeling as-is site condition and tracking the three-dimensional (3D) location of mobile assets (e.g., worker, equipment, material) are essential for various construction applications such as progress monitoring, quality control and safety management. Many efforts have been dedicated to vision-based technologies due to their merits in cost-effectiveness and light infrastructure compared to real-time location systems (RTLS). However, a major challenge of vision-based tracking is that it lacks 3D information and thus the results are sensitive to occlusion, illumination conditions and scale variation. To address this problem, this study presents a point cloud-vision hybrid approach to reconstruct and update the area of interest in 3D for scene updating and mobile asset tracking. Baseline 3D geometry information in point cloud is obtained at the start by Structure from Motion (SfM) using Unmanned Aerial Vehicle (UAV), given which mobile and static assets present in the scene are recognized and labeled. Based on 2D aerial isometric images capture by the UAV, labeled assets are automatically recognized and their locations are updated. The proposed approach was implemented in a field test and the results demonstrate that it was able to reconstruct the site and update the location of mobile assets accurately and reliably. Findings in this study indicate the proposed hybrid approach effectively augments the state-of-the-art in site modeling and asset tracking in construction.

Keywords –

Point cloud-vision hybrid approach; Mobile asset tracking; Structure from Motion (SfM); Unmanned Aerial Vehicle (UAV)

1 Introduction

Three-dimensional (3D) location information of construction assets are of great interest to various engineering and management applications including progress monitoring, quality control, operation analysis, safety monitoring and occupational health assessments. Although much research efforts have been dedicated to investigating the merits of Real-time Location System (RTLS) and vision-based methods, limitations in cost-effectiveness, ease of use, and robustness greatly hinder their field deployment. This study focuses on addressing the challenges in obtaining 3D location data of construction assets through mobile camera systems.

Traditional vision-based tracking methods recognize and track the objects of interest from images or video streams captured by cameras at known locations and angles. Such settings can be easily achieved by installing cameras at multiple locations on a construction site or taking advantage of the existing surveillance camera systems. However, cameras at a fixed location inevitably suffer from massive occlusions introduced by ever-changing site conditions such as structure elements, temporary structures, and equipment. In addition, although 3D location can be computed based on the 2D image captured by two cameras setup at known locations [1], this method requires the tracked objects to be present on both images and not fully occluded. Therefore, continuously tracking the 3D location of construction assets is not always practical at many construction sites with cameras at fixed locations. A mobile camera system on an aerial platform such as an Unmanned Aerial Vehicles (UAV) is considered a promising alternative. Recently, UAV technology has drawn much attention from the construction industry for its potential in various applications including maintenance inspection, construction survey, and safety management. Nevertheless, 3D location tracking using a mobile camera faces several major challenges, including estimating the position and orientation of the on-board camera, and transforming the pixel coordinates of recognized objects from the camera frame to the global frame.

To address these challenges, this paper proposes a point cloud-vision hybrid approach for 3D site reconstruction and mobile asset tracking. This paper first reviews current practices in construction asset tracking and the state-of-the-art Structure from Motion (SfM) technology. Then, the point cloud-vision hybrid approach is introduced by a flowchart and the details of the techniques and algorithms used in each step. Results from a case study implementing the proposed method for vehicle location tracking are presented followed by discussion and conclusions.

2 Related Work

2.1 State-of-the-art Construction Asset Tracking Methods

The location information of construction assets such as workers, equipment, and materials is of interest to various construction applications including progress monitoring, quality control, operation analysis, safety monitoring and occupational health assessments. Much efforts have been dedicated to real-time location systems (RTLS) such as Global Positioning System (GPS) [2], Radio Frequency Identification (RFID) [3], and Ultra-wide Band (UWB) [4]. Although varying in tracking accuracy (e.g., meters for GPS and centimeters for UWB), the RTLS technologies provide direct measurement of the 3D location of the tracked objects. However, most RTLS systems require tagging the objects to be tracked, which increases the complexity in applications. In addition to the tags, high-accuracy RTLS such as UWB requires heavy infrastructure deployment, for example the installation of a series of antennas around the tracking site. This results in huge investment in time and cost ($140/m2).

Another research direction for construction asset tracking focuses on computer vision technologies that recognize and track the objects of interest from 2D images or video streams captured by cameras. Compared to RTLS-based tracking methods, vision-based tracking does not require additional sensors and tags and thus has the advantages of simple deployment and low cost [5]. Various vision-based tracking algorithms have been studied and tested in construction scenarios. Generally, tracking algorithms can be categorized into kernel-based [6], contour-based [7], and point-based [8] methods. Different in the means to represent object, contour-based methods use contours or silhouettes that enclose the object region, kernel-based methods use the responses of the object region to selected kernels, and point-based methods use a set of feature points detected in the object region [9]. Compared to the other two methods, the point-based method is more robust to illumination variation and occlusions, which commonly occur in outdoor construction environment. Although the images from a single camera provide only 2D pixel coordinates, images from multiple cameras at multiple known and fixed locations provide 3D metric coordinates through camera calibration, pose estimation, and triangulation [1].

2.2 Obtaining 3D Information by Photogrammetry and Structure from Motion (SfM)

Photogrammetry is an image-based technology that reconstructs 3D objects from 2D photographs. This technology extracts 2D input data from photographs and maps them onto a 3D space. Since constructing a 3D model only requires taking images from different angles, using photogrammetry for 3D data acquisition is flexible, cost-effective, and non-invasive to the survey objects. Structure from Motion (SfM) photogrammetry is an emerging technique that was built upon but fundamentally differs from traditional photogrammetry. In SfM approach, the critical parameters such as camera location/orientation and scene geometry are automatically computed without the need of a series of targets with known locations [10]. Instead, these parameters are computed simultaneously using a highly redundant, iterative bundle adjustment procedure, based on a database of features automatically extracted from a set of multiple overlapping images [11].

The results of SfM or traditional photogrammetry are represented by a dense point cloud comprised of millions of points, each of which contains 3D position (XYZ) and color (RGB) data. The point cloud data is useful in various construction applications such as acquiring as-is geometry data for building component modeling [12], construction progress monitoring and control [13] [14], and construction documentation especially for historical structures [15]. Although the point cloud contains comprehensive 3D geometric data of the objects in the scene, it has been a challenge to track objects using a SfM-generated point cloud. This is mainly because in SfM technique, generating a point cloud from a large amount of images takes time, as it requires massive computation capability. Therefore, noticeable delay between the actual and tracked location makes it impractical to track dynamic objects on construction sites using SfM method alone.

2.3 Alternative Monitoring Method using Unmanned Aerial Vehicles (UAV)

Many vision-based tracking methods are based on the assumption of using the images captured by cameras at fixed and known locations. This is convenient since many construction sites are equipped with surveillance camera systems. However, construction sites are usually very congested, which makes it impossible for static cameras to constantly maintain a clear line-of-sight to the objects to be tracked without the occlusions from structure elements, equipment, and materials present on the site. Recent development of Unmanned Aerial Vehicles (UAV) offers a low-cost alternative for construction monitoring applications such as bridge and road assessment [16] [17], earthwork surveying [18], and safety inspection [19] . Compared to fixed site cameras, the onboard camera on a UAV is more flexible in image capture angles and thus less prone to occlusions. It should be noticed that a major limitation of UAV-based monitoring is the limited time for a single flight (around 30 minutes) due to the battery life.

3 Point cloud-vision Hybrid Approach

To address the challenges in vision-based tracking using mobile cameras, this study proposes a point cloud-vision hybrid approach for tracking of mobile construction assets using SfM and UAV technologies. The flowchart for the proposed hybrid approach is shown in Figure 1.

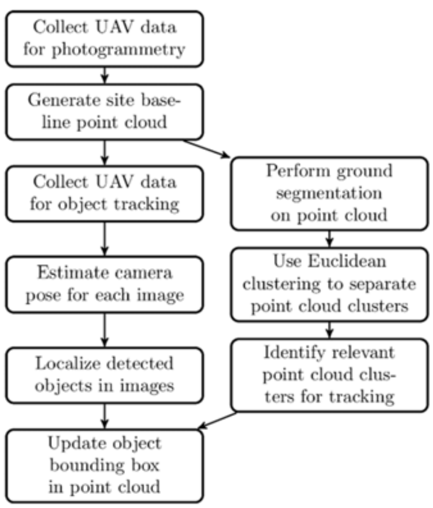

Figure 1: Flowchart for the proposed point cloud-vision hybrid approach

The first step involves collecting aerial images of a construction site from a UAV at multiple viewpoints. The construction site images are highly-overlapping and encircle the site in order to cover the full 3D structure of construction-related entities. Next, the 3D point cloud of the construction site (see Figure 2a) is generated based on the image data using a Structure from Motion (SfM) algorithm adopted from [11]. This algorithm detects common features across each camera frames using Scale Invariant Feature Transform (SIFT). The process works by finding point correspondences between images and solving for point coordinates and camera poses in a bundle adjustment procedure. This establishes a baseline 3D model of the construction site which includes both background elements and the mobile assets. The mobile assets located in the point cloud are separated out using a segmentation and clustering routine. Ground segmentation is first applied to filter out points belonging to the ground which is considered as background. Next, individual point cloud clusters are separated based on neighboring Euclidean distance. Point cloud clusters for objects to be tracked are further identified through a supervised procedure where the user selects a set of clusters corresponding to the interested objects. Bounding boxes are calculated for each identified point cloud cluster which acts as a compact representation of the object.

A series of images is then collected across time from the UAV to specifically track targeted mobile assets. The image data can be either in the form of a video feed or discrete images taken at specific timestamps. Compared to fixed camera setups for object tracking, here it is necessary to solve for the image-specific position and orientation of the camera since the UAV is moving from frame to frame. This is formulated as a perspective-n-point (PnP) problem to recover the complete 6 degree-of-freedom motion (i.e., x, y, z coordinates and yaw, pitch, roll rotations) of the UAV based on the input images. First, an image is synthetically generated from the point cloud projected onto a two-dimensional plane (Figure 2a). Feature points can be calculated from the synthetic image that can be matched to feature points derived from the UAV images. Then a depth buffer is created based on the 3D point cloud data collected from the photogrammetry step. The depth buffer is shown in Figure 2b, where bright points indicate points that are close to the camera while darker points indicate points that are further away from the camera. This enables us to calculate the 3D position of each feature point on the synthetic image. The UAV images are then matched to the synthetic image to obtain a corresponding set of 3D point features. Finally, a camera pose estimate is calculated for each UAV image which minimizes the least-squares re-projection error of 3D point features in the image.

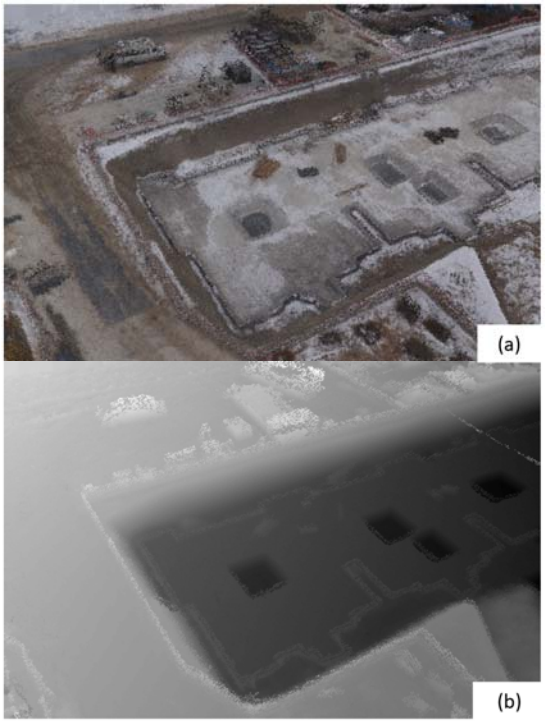

Figure 2: (a) synthetic image and (b) depth buffer generated from point cloud

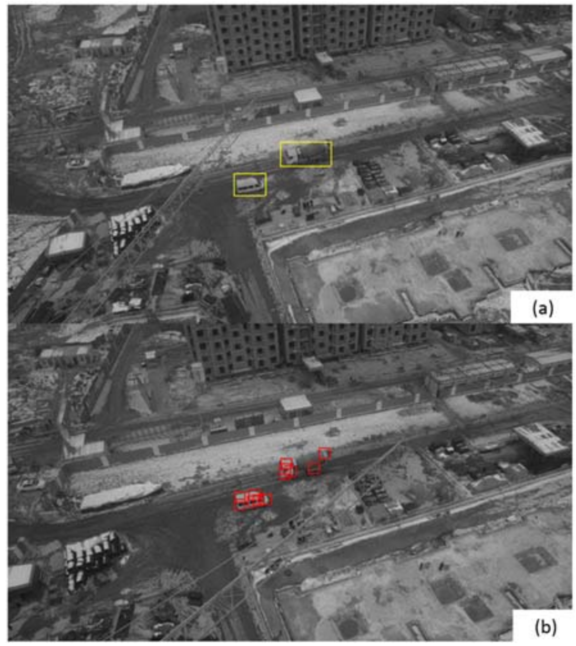

Figure 3: Object detection using feature point matching with a reference image, (a) objects of interest, (b) feature points identified

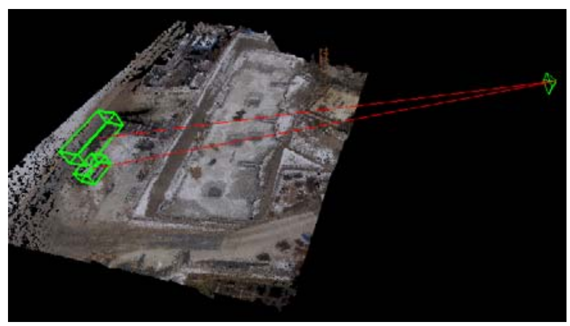

Figure 4: Projection of image coordinates into 3D space from camera origin

The next step in the processing pipeline is object detection. For each image taken from the UAV in the previous step, the pixel coordinates of objects to be tracked are identified through a point-based method, namely matching of SIFT feature points. This process is semi-automated by the user specifying interested objects in the reference image. As shown in Figure 3, bounding boxes are drawn around the two vehicles to be tracked in the reference image (left image) while the same objects are detected in the tracking image through feature point matching (right image). The process can potentially be fully-automated by having a database of possible objects to be tracked or training an object detection classifier.

In the last step, the location of each detected object in global coordinates is calculated based on the recovered camera pose and its image coordinates. A ray casting method is used where the object location is determined by the intersection of a line formed by an image projection vector originating from the camera with the point cloud surface. For each detected object in the image, a corresponding projection vector is determined based on its pixel coordinates and camera parameters such as focal length and image size. Figure 4 shows the projection of detected objects from image coordinates to 3D space. Successfully matched objects have their bounding boxes updated in the point cloud based on the estimated 3D location.

4 Case Study and Preliminary Results

An ongoing construction project was selected as a case study of the proposed method. Two vehicles (i.e., a concrete mixer truck and a minivan) were chosen as the targeted mobile assets to be tracked. Both vehicles moved in random patterns in an area of 40 m by 20 m. This case study employed an 8-axis UAV (octocopter) equipped with a mirrorless digital camera. In total 169 images at a resolution of 4912 x 3264 were used to generate the 3D point cloud of the site. Table x shows the computation time involved in each step of the proposed pipeline. The step of SfM computation and point cloud generation takes the longest amount of time but only needs to be carried out once at the start of the experiment. On the other hand, the tracking and location updating step involves a trade-off between accuracy and computation time. Using high resolution images for tracking will potentially improve the tracking accuracy since more feature points can be detected but this will also increase the computation time.

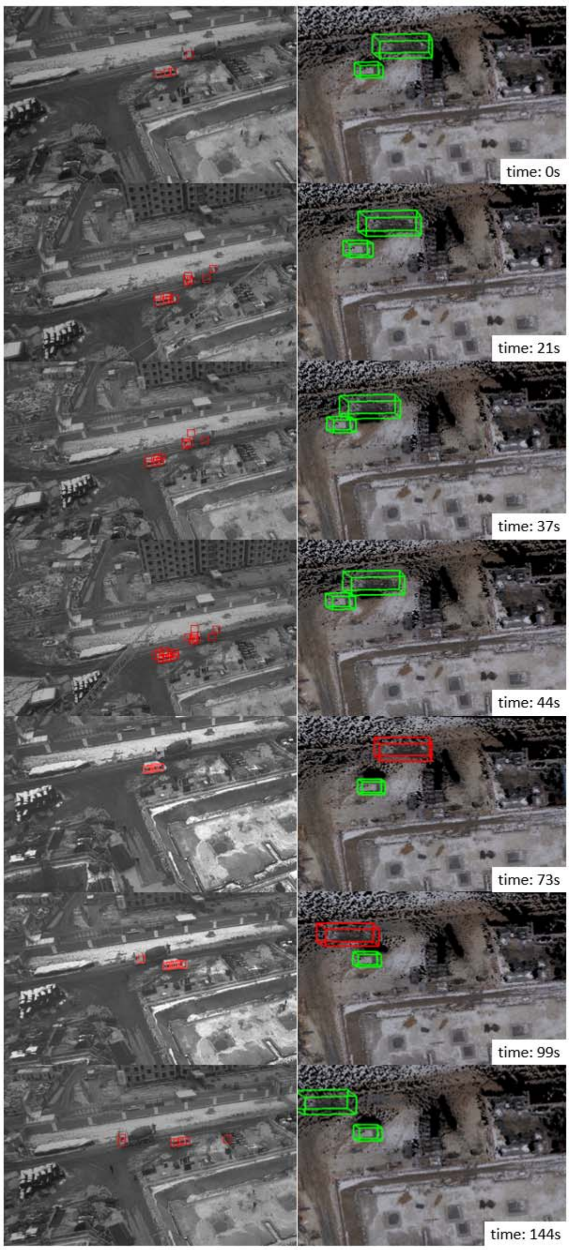

Figure 5: Results of 3D location tracking of two vehicles: image captured from UAV (left) and updated 3D site model (right)

Table 1: Preparation and computation time

Activities	Time
Preparation	~20 min
UAV flight	~3.5 min
SfM computation/Point cloud generation	~150 min
Tracking image by UAV	Up to 30 min
3D location computation	3 min

Figure 5 shows the results obtained from tracking two vehicles over time using images captured from the UAV. The left column shows the time where each image is taken whereas the middle column shows the image captured by the camera. The images are annotated with feature points for each detected object. The right column shows the updated 3D site model (point cloud) corresponding to each captured image. The 3D site model is generated using a top-down view of the site 3D point cloud with bounding boxes formed around each tracked object.

Results from the case study indicate that the proposed method dynamically updates the 3D location of two vehicles in a construction site by using the images captured from a UAV and matching them to a 3D site model in the form of a point cloud. The first four and the last images (time: 0s to 44s and 144s) show the cases of successful tracking where the bounding box for the two vehicles are shown in green. The results at time 73s and 99s show a sequence when the matching algorithm lost track of a vehicle due to insufficient feature points. The corresponding bounding box is drawn based on the previous location estimate but is highlighted in red to indicate uncertainty in location.

5 Discussion

In this study, the tracked objects focuses on large objects exhibiting smooth linear motion such as vehicles. Smaller objects that are prone to occlusion incur difficulty in the feature point matching stage. The method is also limited in terms of automation in the sense that not all dynamic assets are automatically tracked and updated in the 3D point cloud. Instead, the user manually specifies a fixed number of objects to be tracked in the 3D scene update.

In terms of accuracy in location estimation, the method relies on accurately calculating the UAV camera position and orientation for each captured image and correctly identifying the tracked object in each image. The camera pose is derived through a least-squares estimation scheme and is negatively affected by the presence of outliers. In the matching stage with a reference image, outlier points have to be rejected by threshold elimination considering the re-projection error to the image coordinate frame. In the object detection stage, the pixel coordinates of tracked objects can also be incorrectly identified when there exists feature points that are incorrectly matched. Thus, the matching algorithm needs to ensure that a sufficient number of feature points can be identified for each object and that there exists geometric consistency among the matched feature points for objects from each frame to the next frame.

6 CONCLUSION

To address the challenges in 3D location tracking using a mobile camera, this paper proposes a point cloud-vision hybrid approach for 3D site reconstruction and mobile asset tracking. The method involves a processing pipeline with the steps of point cloud generation, camera pose estimation, object detection and object localization. The method updates the baseline 3D scene in the form of point cloud with dynamic bounding boxes for each tracked vehicle, which can be further utilized in site management applications. Preliminary results from a case study show that the proposed method was able to successfully track the 3D location of two vehicles in a construction site by using images captured from a UAV and matching them to a 3D site model in the form of a point cloud. Despite of the limitations in semi-automated processing pipeline and limited UAV flight time, the proposed point cloud-vision hybrid approach enable by SfM and UAV technology shows advantages in flexibility and robustness to occlusions over traditional method using fixed cameras. Findings in this study indicate great potential of the proposed method in 3D location tracking of mobile construction assets in congested construction environment.

The proposed method involves limitations in terms of the number and size of objects that can be tracked. Having a large number of tracked objects or having target objects that are too small will complicate the feature point matching process and reduce the localization accuracy. For future work, the authors would like to experiment with alternative vision tracking methods such as kernel-based and contour-based methods to investigate whether those methods will improve the localization accuracy.

7 ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation (Award #: CMMI-1358176). Any opinions, findings, and conclusions or recommendations expressed on this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] M.-W. Park, C. Koch, and I. Brilakis, “Three-Dimensional Tracking of Construction Resources Using an On-Site Camera System,” J. Comput. Civ. Eng., vol. 26, no. 4, pp. 541–549, 2012.

[2] N. Pradhananga and J. Teizer, “Automatic spatio-temporal analysis of construction site equipment operations using GPS data,” Autom. Constr., vol. 29, pp. 107–122, 2013.

[3] Y. Fang, Y. K. Cho, S. Zhang, and E. Perez, “Case Study of BIM and Cloud–Enabled Real-Time RFID Indoor Localization for Construction Management Applications,” J. Constr. Eng. Manag., pp. 1–12, 2016.

[4] Y. K. Cho, J. H. Youn, and D. Martinez, “Error modeling for an untethered ultra-wideband system for construction indoor asset tracking,” Autom. Constr., vol. 19, no. 1, pp. 43–54, 2010.

[5] J. Yang, M.-W. Park, P. a. Vela, and M. Golparvar-Fard, “Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future,” Adv. Eng. Informatics, vol. 29, no. 2, pp. 211–224, 2015.

[6] E. Maggio and A. Cavallaro, “Accurate appearance-based Bayesian tracking for maneuvering targets,” Comput. Vis. Image Underst., vol. 113, no. 4, pp. 544–555, 2009.

[7] M. Yokoyama and T. Poggio, “A contour-based moving object detection and tracking,” 2005 IEEE Int. Work. Vis. Surveill. Perform. Eval. Track. Surveill., no. 1, pp. 271–276, 2005.

[8] T. Mathes and J. Piater, “Robust non-rigid object tracking using point distribution manifolds,” Pattern Recognit., pp. 515–524, 2006.

[9] I. Brilakis, M.-W. Park, and G. Jog, “Automated vision tracking of project related entities,” Adv. Eng. Informatics, vol. 25, no. 4, pp. 713–724, Oct. 2011.

[10] M. J. Westoby, J. Brasington, N. F. Glasser, M. J. Hambrey, and J. M. Reynolds, “‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications,” Geomorphology, vol. 179, pp. 300–314, 2012.

[11] N. Snavely, S. M. Seitz, and R. Szeliski, “Modeling the world from Internet photo collections,” Int. J. Comput. Vis., vol. 80, no. 2, pp. 189–210, 2008.

[12] F. Dai and M. Lu, “Assessing the Accuracy of Applying Photogrammetry to Take Geometric Measurements on Building Products,” J. Constr. Eng. Manag., vol. 136, no. February, pp. 242–250, 2010.

[13] S. El-Omari and O. Moselhi, “Integrating 3D laser scanning and photogrammetry for progress measurement of construction work,” Autom. Constr., vol. 18, no. 1, pp. 1–9, 2008.

[14] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, and F. Peña-Mora, “Evaluation of image-based

modeling and laser scanning accuracy for emerging automated performance monitoring techniques," Autom. Constr., vol. 20, no. 8, pp. 1143–1155, 2011.

[15] N. Yastikli, "Documentation of cultural heritage using digital photogrammetry and laser scanning," J. Cult. Herit., vol. 8, no. 4, pp. 423–427, 2007.

[16] N. Metni and T. Hamel, "A UAV for bridge inspection: Visual servoing control law with orientation limits," Autom. Constr., vol. 17, no. 1, pp. 3–10, 2007.

[17] S. Rathinam, Z. W. Kim, and R. Sengupta, "Vision-Based Monitoring of Locally Linear Structures Using an Unmanned Aerial Vehicle," J. Infrastruct. Syst., vol. 14, no. 1, pp. 52–63, 2008.

[18] S. Siebert and J. Teizer, "Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system," Autom. Constr., vol. 41, pp. 1–14, 2014.

[19] J. Irizarry, M. Gheisari, and B. N. Walker, "Usability assessment of drone technology as safety inspection tools," Electron. J. Inf. Technol. Constr., vol. 17, no. September, pp. 194–212, 2012.