AUTOMATIC OBJECT RECOGNITION AND REGISTRATION OF DYNAMIC CONSTRUCTION EQUIPMENT FROM A 3D POINT CLOUD

Mengmeng Gai and *Yong Cho
University of Nebraska – Lincoln
1110 S. 67th St. Peter Kiewit Institute (PKI) 118A,
Omaha, NE, USA 68182-0571
*(*Corresponding author: ycho2@unl.edu)*

A FRAMEWORK FOR AUTOMATIC OBJECT RECOGNITION AND REGISTRATION OF DYNAMIC CONSTRUCTION EQUIPMENT FROM A 3D POINT CLOUD

ABSTRACT

This paper introduces a model-based automatic object recognition and registration framework to assist heavy equipment operators in rapidly perceiving 3D working environment at dynamic construction sites. A video camera and a laser scanner were utilized in this study to rapidly recognize and register dynamic target objects in a 3D space by dynamically separating target object's point cloud data from a background scene for a quick computing process. A smart scan data updating method has been developed which only updates the dynamic target object's point cloud data while keeping the previously scanned static work environments. Extracted target areas containing 3D point clouds were orthographically projected into a series of 2D planes with a rotation center located in the target's vertical-middle line. Prepared 2D templates were compared to these 2D planes by extracting SURF features. Point cloud bundles of the target were recognized, and followed by the prepared CAD model's registration to the templates. The field experimental results show that the proposed rapid workspace modeling method can significantly improve heavy construction equipment operations and automated equipment control by rapidly modeling dynamic target objects in a 3D view.

KEYWORDS

Equipment, 3D, point cloud, object recognition, construction safety

INTRODUCTION

The interactions between workers, equipment, and materials can easily create visibility-related accidents. Visibility problems can lead to serious collisions without pro-active warnings. Because lack of full visibility is a major contributing factor in accidents at construction sites, there have been a number of advances in vision-aid techniques. 3D spatial modeling can help to optimize equipment control, significantly improve safety (Teizer, Allread, Fullerton, & Hinze, 2010), and enhance a remote operator's spatial perception of the workspace (Cho, Wang, Tang & Haas, 2012). However, the rapid processing of tens of thousands bits of range data in real is still an unsolved problem requiring further investigation (Gong & Caldas, 2008). Unstructured work areas like construction sites are difficult to visualize graphically because they involve highly unpredictable activities and change rapidly. Construction site operations require real-time, or near real-time information, about the surrounding work environment, which further complicates graphical modeling and updating.

One commonly used method to obtain the 3D position of an object is based on 3D scanning technology (Cho et al., 2012; Tang, Huber, Akinci & Lipman, 2010; Huber, Akinci, Tang & Adan, 2010); this method, however, has some limitations, such as low data collection speed and low object recognition rates (Kim, Lee, Cho & Kim, 2011). It has always been a challenge to recognize specific objects from a 3D point cloud in unstructured construction environments because it is difficult to rapidly extract the target area from background noises in a large and complex 3D point clouds.

While rapid workspace modeling is essential to effectively control construction equipment (Lee, et al., 2009), few approaches have been accepted by the construction industry due to the difficulty of addressing all the challenges of current construction material handling tasks with the current sensor technologies. Thus, an innovation in rapid 3D spatial information is necessary to meet the challenges. Based on the previous work (Gai, Cho & Wang, 2012), the main objective of this research is to propose a model-based automatic target object recognition and registration method to help heavy equipment operators rapidly perceive 3D working environments at dynamic construction sites.

This paper is organized as follows. First, a literature review of state-of-the-art object tracking and visualizing techniques used on construction job sites will be given. Then, a model-based framework will be discussed. After that, the validation of construction field test will be presented, and followed by conclusions and future work.

LITERATURE REVIEW

This section mainly discusses the state-of-the-art object tracking and visualizing techniques for real-time applications on construction sites.

Sensor-based Methods: In the early stage, Radio Frequency Identification (RFID) and ultra-wideband were adopted in tag-based system to detect moving objects. Global Positioning System (GPS) and web-based technologies were implemented to track vehicles and detect collision at outdoor environments (Navon & Shpatnitsky, 2005). There are also some attempts to combine RFID with GPS technology, and transfer data between detectors and receivers (Andoh, Su & Cai, 2012). However, GPS has drawbacks such that it works ineffectively without direct line of sight from the satellites, and it is expensive to install on every moving object, or parts of equipment. RFID readers need to be equipped in the equipment and connected to the computer networks for exchange of information, which mean additional costs related to hiring additional hardware and technical consultants.

Vision-based Methods: A camera system consisting of one camera on the rear axle of the truck and one camera on the front of truck, as well as a video monitor in the cab can provide a visual check of the front and rear blind areas (Ruff 2007). Stereo vision-based methods, based on computer vision technologies, have been proposed as an effective alternative for tracking moving entities. Brilakis, Park and Jog (2011) introduced 2D vision-based methods that recognize new overlapping feature points and track them in subsequent video stream. To acquire a precise 3D position of objects with additional depth information, two or more cameras generate a stereo view after calibration with known intrinsic parameters. Park et al. (2012) achieved more accurate 3D locations of tracking objects by projecting the centroids of two cameras to 3D coordinates. There are two known drawbacks of vision-based techniques in tracking moving equipment at sites: 1) fixed camera locations have limited view angles and resolutions, and 2) the results are sensitive to lighting conditions.

Laser Scanner-based Methods: Laser scanners have been extensively utilized to automatically obtain the "as-built" condition of an existing building; they also can be used to classify and capture a complex heavy equipment operation as it happens or to provide automating feedback to those who are conducting the operations (Arayici, 2007; Gai et al., 2012; Cho et al., 2012). Lee et al. (2009) proposed an automated lifting-path tracking system on a tower crane to receive and record data from a laser device. Teizer et al. (2010) used a laser scanner inside an equipment cab to detect blind spots from 3D point clouds. Bosche and Hass (2008) registered 3D static CAD objects to laser-scanned point cloud data, which can be utilized to efficiently assess construction processes. However, most of the algorithms were developed mainly to recognize and register static objects' models to point clouds. Few applications have demonstrated the technical feasibility of registering dynamic models to point clouds in real time or near real time.

METHODS

A model-based rapid automatic object recognition and registration framework has been developed to help heavy equipment operators rapidly perceive the 3D working environment at dynamic construction sites. The framework of the proposed method is shown in Figure 1.

This framework is mainly composed of main components: images and point cloud correlation (IPCC), video-based object recognition and tracking (VORT), target-focused smart scanning and updating (TSSU), 3D-2D projection and object recognition (3POR) and results visualization. In IPCC, a correlation between image data and 3D point cloud is built; as a result, images are mapped onto the point clouds. The following components will be implemented based on the mapped point cloud data. In VORT, a calibrated video camera provides 2D consequential images of moving objects. Equipment operators define single or multiple bounding areas that contain specific single or multiple moving targets − like materials or equipment − from the graphical user interface (GUI). The Speeded Up Robust Features (SURF) features are extracted from the selected bounding areas. 2D consequential images with extracted SURF features are simultaneously compared to the features of the selected bounding areas. As a result, single or multiple 2D target areas are stretched and updated in the images. Taking a yellow robot as an example, Figure 2 shows the recognition and registration process. As a target, the yellow robot is automatically scanned and updated. Corresponding to the 2D target regions obtained in the last component, 3D bounding areas are scanned in the following rounds to replace the previously scanned work environments. Extracted target areas containing 3D point clouds are projected to a series of 2D planes with a rotation center located in the targets' vertical-middle line. By extracting their SURF features, prepared 2D templates, which contain the operator's selected objects, are compared with 2D planes.

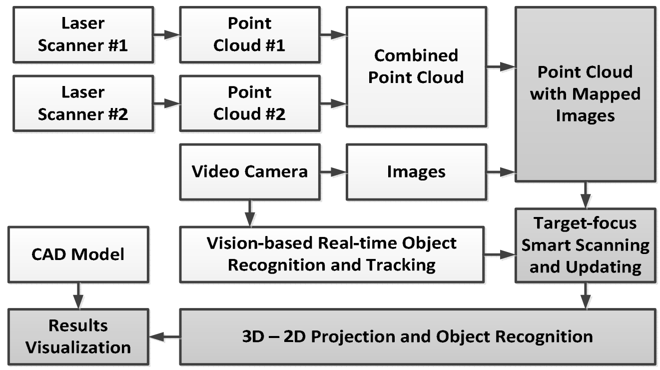


Figure 1. The framework of the proposed method

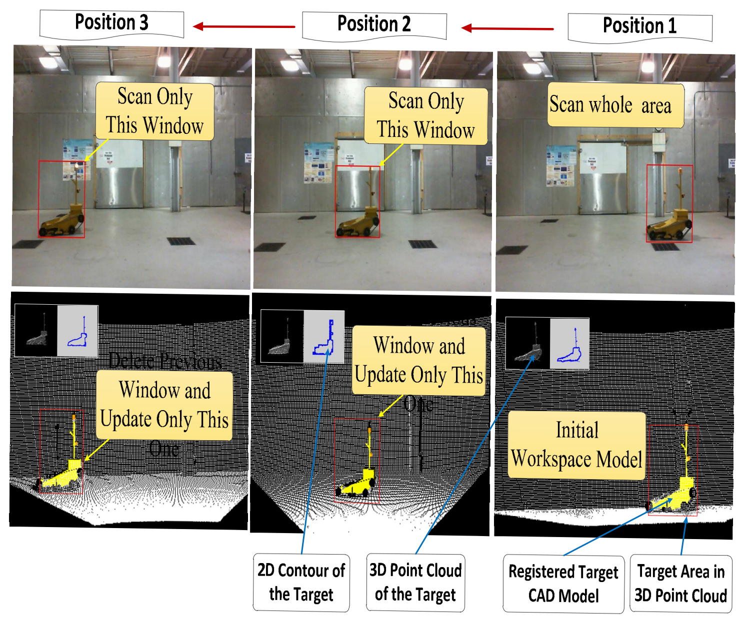


Figure 2. Example of object recognition and registration process

Data Acquisition: An innovative robotic hybrid Light Detection And Ranging (LIDAR) system was developed, consisting of two 2D line laser scanners (80 meter working ranges at 100Hz scan speed, up to 2.5 sec / 360º scan, 190º for vertical line), and a video camera, as shown in Figure 3. The resolution of each line laser is 0.25 degree in a vertical direction and 0.0072 degree in a horizontal direction. The customized 3D LIDAR system provides more flexibility in hardware control and software programming than a commercial LIDAR scanner does (Gai et al., 2012; Cho & Martinez, 2009). Based on the mounting configuration, we solved multiple degree-of-freedom (DOF) kinematics to obtain x-y-z coordinates from the LIDAR, and simultaneously generated real-time digital image data from the video camera. The LIDAR system was designed with two reversely positioned 2D line scanners to provide twice faster scanning rate and twice higher scanning resolution (Figure 4).

Target Area-focused Smart Scanning: The component of target area-focused smart scanning is implemented based on the 2D bounding area obtained in 2D image-based object recognition and tracking. Based on the bounding box approach, smart scanning can reduce data size and scanning time (Cho et al., 2012). A video camera captures 2D consequential images of working environments containing different types of equipment. The SURF descriptor was utilized for visual object recognition, and based on the results of which the algorithm Kalman filter (Steffen, 1981; Steffen, 2002) was used in the object tracking phase to produce estimates of unknown variables that are more precise than those based on a single measurement. The bounding box (i.e., Kernel) areas, containing specific moving objects (i.e., materials, a whole piece of equipment, or parts of heavy equipment), were simultaneously defined and stored through the developed GUI as templates from which SURF features were extracted. 2D consequential images, from which SURF features were extracted, were provided by the video camera and compared to the templates, producing the common SURF features. As a result, multiple 2D target bounding box areas were defined based on the Recognition Qualification Value (RQV) from the SURF descriptors and updated from the images. Also, the target areas were used to update the template data set and used for the laser scanning. Namely, laser scanners only update the dynamic target object's point cloud data while keeping the previously scanned static work environments.

Figure 3. LIDAR system configuration

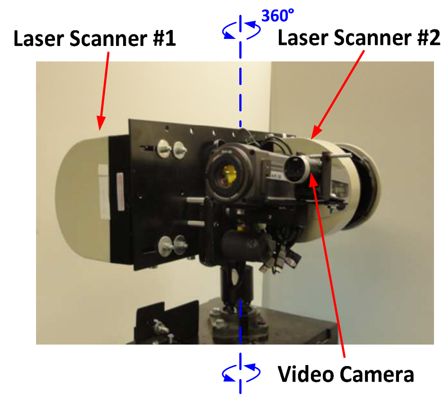


Figure 4. Prototype LIDAR system

Object Recognition and Registration: Orthographical projection from 3D to 2D, a process of mapping a 3D point cloud to a 2D plane, is introduced to recognize and localize the target in a 3D view. Gathered by the hybrid laser system, the 3D point cloud is orthographically projected into different 2D planes from different directions. Assuming that point O (O_x, O_y, O_z) is orthographically projected onto 2D points O_1(O_{1x}, O_{1z}) parallel to the y axis, the coordinate values of point O1 can be calculated as follows (Equation 1):

$$\begin{bmatrix} O_{1x} \\ O_{1z} \end{bmatrix} = \begin{bmatrix} m_x & 0 & 0 \\ 0 & 0 & m_z \end{bmatrix} \begin{bmatrix} O_x \\ O_y \\ O_z \end{bmatrix} + \begin{bmatrix} n_x \\ n_z \end{bmatrix} \tag{1}$$

where m is an arbitrary scale factor and n is an arbitrary offset factor, both of which can be used to align the projection viewport. Offline templates of the targets are prepared based on different projection angles, then stored in the local software database. Individual templates are generated one by one and common features are extracted from the corresponding shapes. Projected 2D planes as input of the target recognition component are processed online and corresponding shapes and common features are generated from them. Similarity comparison between common features from projected 2D planes and those from the templates database is then implemented. Finally, comparison qualification values are generated, from which the template corresponding to the minimum one is chosen as the process result.

The main challenge from the process of object recognition in this study is to extract the shapes from projected 2D planes and compare the corresponding shapes with the prepared templates in the database. A local descriptor SURF (Bay, Tuytelaars & Gool, 2008) and the methodology process provided by Mikolajczyk et al. (2003) were employed to perform the target shape recognition from 2D planes. The whole system is composed of two main stages: 1) reducing ambiguity via a local transformation and 2) implementing object detection by estimating a global transformation (Mikolajczyk et al., 2003).

Prepared CAD models correspond one-to-one to the prepared point cloud templates (Figure 5). A series of 2D planes are projected from extracted target areas, from which object contours are extracted, followed by a filtering process to remove the outliers from the corresponding SURF features. In order to filter the extracted features contaminated by outliers, several methods have been proposed with promising results, such as the Random Sample Consensus (RANSAC) algorithm (Fisher & Bolles, 1981). In this study, a triangle relationship-based filtering method is used to remove the outliers from the feature data array.

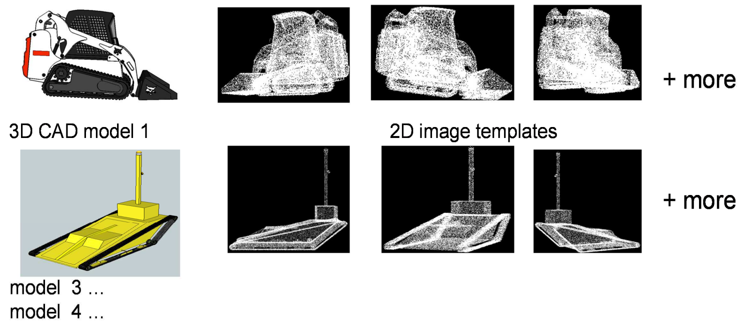


Figure 5. Prepared CAD models and generated point cloud templates

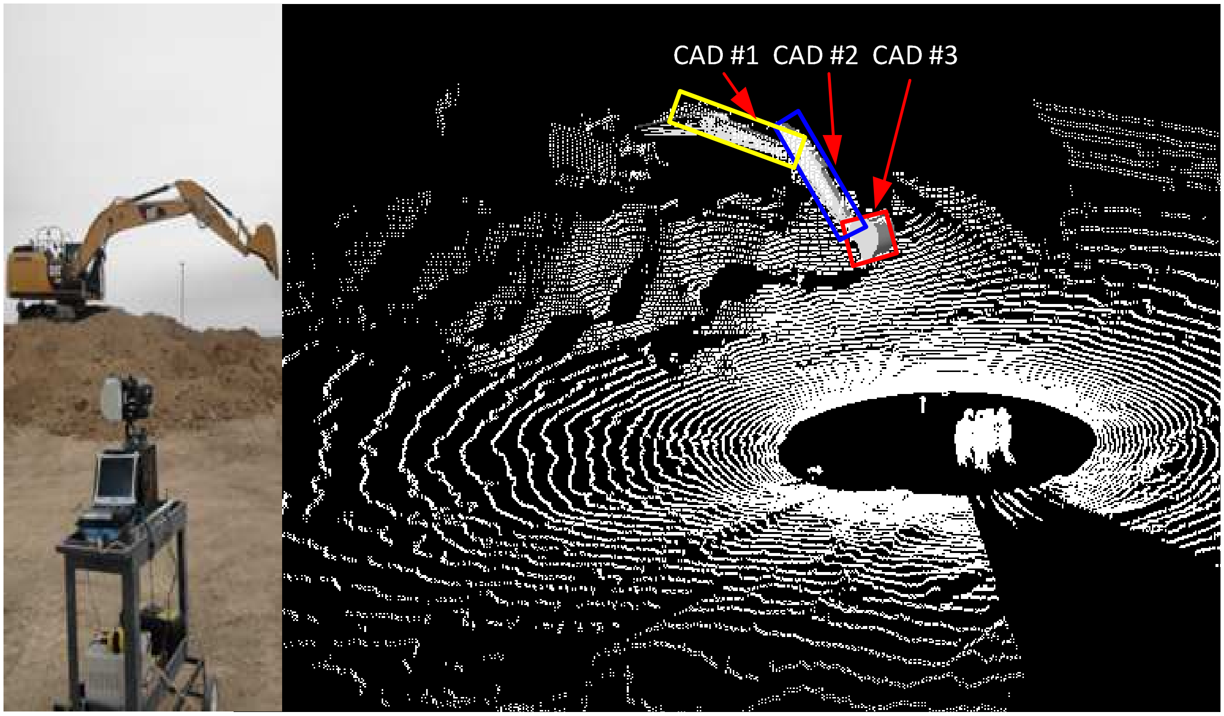
The output data array performs a reverse calculation of 3D point cloud to 2D planes after the outliers are removed from the original contaminated features. 3D position calculation is the reverse projection process from 3D point clouds to 2D planes. The object is located in the projection center of the coordinate system, and different projection angles can be randomly selected from the laser scanning direction. Based on the 3D coordinate value of the object contour, an existing CAD model from a database that has same dimension as the object is aligned according to the coordinate values of the object in a 3D view.

RESULTS

In this study, an earthmoving construction equipment operation site was visited to validate the proposed framework. Mounted on a mobile cart (Figure 6a), the hybrid LIDAR system gathered point clouds with digital images of the job sites that contained different types of heavy equipment and their working spaces. In the job site field tests, a separate data server connected to the hybrid laser system was designed to automatically store the scanned data set from dynamic working environments and share the data through a wireless communication technology. Multiple equipment operators can access the data produced by the LIDAR system via mobile terminals and investigate the real-time situation of the surroundings. Figure 6 (b) gives an example of the point cloud data with rapidly registered CAD models of three key moving target parts of the excavator: boom, arm, and bucket. In this experiment, the data transferring speed between the LIDAR system and terminals was around 25M/bps, which was fast enough to update the scene in real time. The excavator operator selected three independent parts of the equipment as targets. Then the developed system continuously provided time-elapsed 3D scenes of the whole working environment with the registered CAD models to help the operator rapidly perceive the 3D working environment from different viewpoints.

CONCLUSIONS

In this study, a framework for rapid workspace modeling for construction equipment operations was introduced to help heavy equipment operators instantly perceive the 3D working environment at dynamic construction sites, and to improve construction equipment operation safety and productivity. This method significantly reduces the size of data to collect and process and improves computing efficiency while keeping the surrounding spatial information. Field demonstrations were successfully conducted to validate the technical feasibility of the proposed framework and hardware systems. Vision-based rapid object recognition and tracking was implemented and 3D data were transferred through a local network in real time.



(a) (b)

Figure 6. Example of the point cloud data with registered CAD models shown in the operator's screen: (a) hybrid LIDAR system and test environment; (b) Point cloud and registered three CAD models

For future work, the research will continue to improve the resolution of LIDAR data while reducing data collection time. With an increase in scanning speed, the scanned resolution is lowered accordingly. To resolve this issue, a smart scanning approach with differentiated scan speeds will be further developed, to allow faster rotations for the areas to be skipped, and slow the scan speed for the target areas. Another alternative is to add one or two more line lasers to the current system, which will significantly increase the scan rate.

REFERENCES

Andoh, H. A., Su, X., and Cai, H. (2012). "A framework of RFID and GPS for tracking construction site dynamics." Construction Research Congress 2012, ASCE, 818-827.

Arayici, Y. (2007). "An approach for real world data modeling with the 3D terrestrial laser scanner for built environment." Journal of Automation in Construction, 16 (6), 816–829.

Bay, H., Tuytelaars, T., and Gool, L. V. (2008). "SURF: Speeded up robust features." Computer Vision–ECCV, 404-417.

Bosche, F., and Haas, C. T. (2008). "Automated retrieval of 3D CAD model objects in construction range images." Journal of Automation in Construction, 17(4), 499-512.

Brilakis, I., Park, M., and Jog, G. (2011). "Automated vision tracking of project related entities." Advanced Engineering Informatics, 25, 713-724.

Cho, Y., and Martinez, D. (2009). "Light-weight 3D LADAR system for construction robotic operations." 26th International Symposium on Automation and Robotics in Construction (ISARC), Austin, Texas.

Cho, Y., Wang, C., Tang, P., and Haas, C. (2012). "Target-focused local workspace modeling for construction automation applications." ASCE Journal of Computing in Civil Engineering, 26(5). 661-670.

Fisher, M., and Bolles, R. (1981). "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography." Comm. of the ACM, 24(6), 381–395.

Gai, M., Cho, Y., and Wang, C. (2012). "Projection-Recognition-Projection (PRP) method for object recognition from a 3D point cloud." Proceedings of the 2012 ASCE International Workshop on Computing in Civil Engineering, Clearwater, FL.

Gong, J., and Caldas, C. H. (2008). "Data processing for real-time construction site spatial modeling." International Journal of Automation in Construction, 17 (5), 526-535.

Huber, D., Akinci, B., Tang, P., Adan, A., et al. (2010). "Using laser scanner for modeling and analysis in architecture, engineering and construction," Proceedings of the Conference on Information Sciences and Systems (CISS), Princeton, NJ.

Kim, C., Lee, J., Cho, M., and Kim, C. (2011). "Fully automated registration of 3D CAD model with point cloud from construction site," 28th International Symposium on Automation and Robotics in Construction, Seoul, Korea, 917-922.

Lee, G., Kim, H., Lee, C., Ham, S., Yun, S., Cho, H., Kim, B., Kim, G., and Kim, K. (2009). "A laser-technology-based lifting-path tracking system for a robotic tower crane." Automation in Construction. 18(7), 865-874.

Mikolajczyk, K., Zisserman, A., Schmid, C. (2003) "Shape recognition with edge-based features." Proceedings of the British Machine Vision Conference, Norwich, UK.

Park, M., Koch, C., and Brilakis, I. (2012). "Three-dimensional tracking of construction resources using an on-site camera system." J. Comput. Civ. Eng., 26(4), 541-549.

Ruff, T.M. (2007). "Recommendations for evaluating & implementing proximity warning systems on surface mining equipment." Research Report to the National Institute for Occupational Safety and Health, Centers for Disease Control, < http://www.cdc.gov/niosh/mining/works/coversheet202.html >.

Steffen L. Lauritzen, Thiele: Pioneer in Statistics, Oxford University Press, 2002. ISBN 0-19-850972-3.

Tang, P., Huber, D., Akinci, B., Lipman, R. et al. (2010). "Automatic reconstruction of as-build building information models from laser-scanned point clouds: A review of related techniques." Automation in Construction, 19, 829-843.

Teizer, J., Allread, B. S., Fullerton, C. E., and Hinze, J. (2010). "Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system." Automation in Construction, 19(5), 630-640.

Teizer, J., Allread, B. S., Mantripragada, U. (2010). "Automating the blind spot measurement of construction equipment." Automation in Construction, 19 (4), 491-501.