Key Nodes Modeling for Object Detection and Location on Construction site using Color-Depth Cameras

Chenxi Yuan[1] and Hubo Cai[2]

[1]Graduate student, Division of Construction Engineering & Management, School of Civil Engineering, Purdue University, 550 Stadium Mall Dr., West Lafayette, IN 47907; email: yuan89@purdue.edu
[2]Assistant Professor, Division of Construction Engineering & Management, School of Civil Engineering, Purdue University, 550 Stadium Mall Dr., West Lafayette, IN 47907, PH (765) 494-5028; email: hubocai@purdue.edu

ABSTRACT

Object detection and tracking is a challenging problem in the dynamic construction environment, especially for objects with kinematic joints and changing poses, e.g. excavators. This paper presents a kinematic key nodes model for dynamically detecting and locating movable objects in construction sites using color-depth cameras. The key nodes model is designed by first analyzing kinematic constraints from mechanic specifications of interested construction site objects and then building representative segmented contour templates under supervision. Feature matching is realized by comparing the real object components and segmented contour templates in a traversal order of the tree structure for all the key nodes. Since color-depth cameras capture not only images, but also the distance between detected objects and the camera itself in a format of three-dimensional (3D) point clouds, the 3D spatial location of an object is obtained by linking the image coordinate of each detected key node with the 3D point coordinates. This proposed framework showed great potential in construction monitoring and safety management.

INTRODUCTION

Construction industry accounts for the highest number of fatal work injuries among all industry sectors in 2012 (OSHA 2013). One-fourth of the construction worker deaths are caused by equipment-involved incidents (Hinze and Teizer 2011). Invisibility issue such as “blind spot” and “blind lift” is an important reason inducing equipment related fatalities in such a dynamic and cluttered environment (Gai et al. 2013). Therefore, how to detect and locate the movable equipment in real-time or near real-time is a pivotal issue to improve safety for construction site workers.

Although construction equipment have installed several safety devices such as horns, head lights, brake lights, directional signals, mirrors, etc. (MacCollum 1995), the warning signal is a passive safety precaution which works only if the passing-by workers notice it. In order to actively monitor and detect equipment, vision-based digital cameras and camcorders have been used on construction site. For example, (Rezazadeh Azar and McCabe 2012) used computer vision algorithms to automatically detect excavators and dump trucks. Stereo imaging is a three dimensional technique using two or more 2D cameras to calculate depth value, but it is not an easy task for real time detection owing to the limitations of the relevant algorithms (Gai and Xu 2013; Gore et al. 2011; Uslu et al. 2011). 3D Laser Scanner is quite a new range sensing technology with high accuracy. Given its low data collection speed and low object recognition rates (Kim et al. 2011), it is only worthwhile to do three-dimensional as-built reconstructions at some important construction time nodes (Argüelles-Fraga et al. 2013; Bosché 2012; Tang and Alaswad 2012). Consumer depth camera has emerged as an inexpensive range sensing technology that combines the merits of both 2D cameras and 3D laser scanners. It captures both images and three-dimensional point clouds simultaneously at a relatively high video frame rate, which is ideal to detect movable objects on construction sites in a sequential manner. Therefore, in this paper, consumer depth cameras will be used as the data acquisition sensor.

Once data is collected, the main concern becomes how to detect objects efficiently and effectively. A number of researchers have investigated the potential of computer vision algorithms to detect and monitor construction objects. For instance, image color space was used to detect hydraulic excavator by Zou and Kim (2007). Histogram of Oriented Gradients (HOG) was used to train templates and detect dump trucks by Rezazadeh Azar and McCabe (2012). However, object detection and tracking is still a challenging problem in construction in the following two aspects: 1) Construction site is a dynamic and scattered environment, it is difficult to segment object out of surroundings; 2) it is a time-consuming and complex process to train templates for equipment with kinematic joints and changing poses, e.g. excavators.

In order to solve the above challenges, this paper presents a kinematic key nodes model for dynamically detecting and locating movable objects in construction sites using color-depth cameras. The remainder of the paper is organized as follows. First, key contributions are highlighted. Following the contributions section, technical details regarding the image training and kinematic features extraction, key node models structure design and the feature matching process are discussed in the methodology section. Finally, the authors discuss research findings, draw conclusions, and point out future research directions.

CONTRIBUTIONS

The kinematic key nodes model proposed in this study has two merits. First, it can be used as shape descriptors for deformable object feature extraction and template training. It extracts semantic features from mechanic specification instead of blindly relying on unsupervised training and estimation. Therefore, a more flexible and structural features can be obtained and meanwhile a lot of efforts can be avoided on optimizing complex training algorithms.

Second, color-depth camera can capture synchronized image pixels and point cloud, and no extra effort or additional processing is needed to align resulting 2D images and 3D point clouds. Therefore, after objects are detected and segmented from the images, the distances between the color-depth camera and the interested key nodes can be directly obtained. Then the analysis of interactions between objects can be easily achieved by merely calculating the distance between several key nodes.

METHODOLOGY

The main objective of this study is to present a new method on object detection and location based on the proposed key nodes model using color-depth cameras. This method includes following four main tasks (see Fig.1).

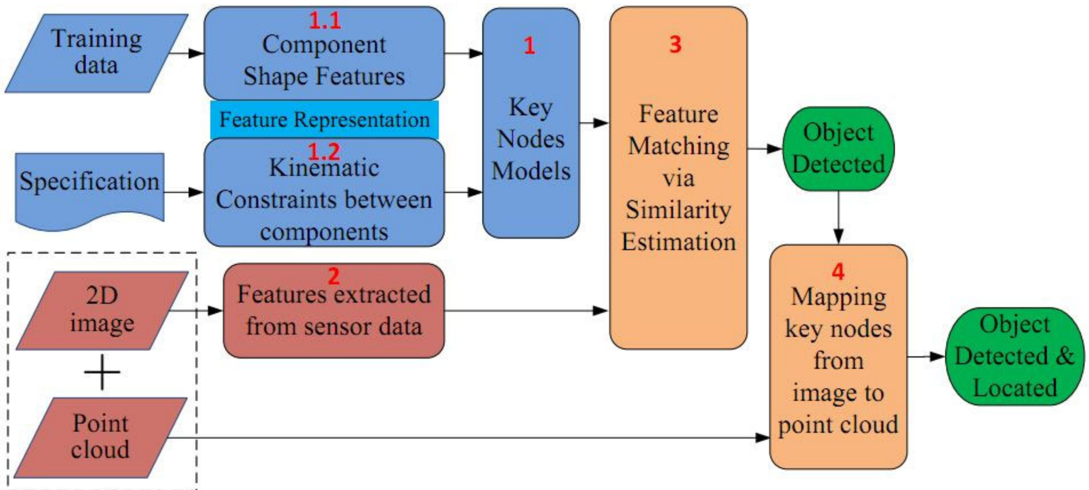


Figure 1. Four tasks for object detection and location

1) **The training process** designed for this study is to extract the geometrical features of objects from training data and meanwhile to incorporate the kinematic constraints according to the mechanic specification.

1.1) The training data is collected from several representative 2D images in different view angles. The boundary of the object is extracted by first using Canny Edge Detector (OpenCV2.4.7.0) and then manually segmenting each component from the background.

1.2) The kinematic constraints between adjacent components, which can be obtained from the mechanic specification, are incorporated into the key nodes model. After multiple combinations, more contour templates can be derived under supervision based on the structure of the key nodes model.

2) **Feature Extraction from 2D image of the sensor data**: The task in this step is to compute an edge map for the 2D image captured by the color-depth camera. In this study, the Canny Edge Detector(OpenCV2.4.7.0) is also used.

3) **Feature matching** is realized by comparing the real object components captured from 2D image in step2 and segmented contour templates in step1 according to a traversal order of the tree structure for all the key nodes.

4) **2D to 3D mapping**: The relative spatial location of the detected object is obtained by mapping the image pixels of detected key nodes into 3D coordinates of the point cloud.

In the following sub-sections, the specification of CAT 336D L hydraulic excavator (CAT) will be investigated as an example, and then an excavator toy model will be used to demonstrate the concept of the new methodology.

Canny Edge Detector. Canny edge detector is a robust edge detection operator developed by John F. (Canny (1986)) which uses a four-steps algorithm to detect the edge of images: 1) smooth image with a Gaussian kernel to filter out noise; 2) compute the gradient magnitude of the image; 3) keep only thin edges or lines; 4) detect edges by lower and upper thresholds. Figure 2b shows the edge extraction result of an excavator toy model using Canny detector. All the important edges of the excavator have been kept. Figure 2c shows the binary image of an excavator after manually segmentation.

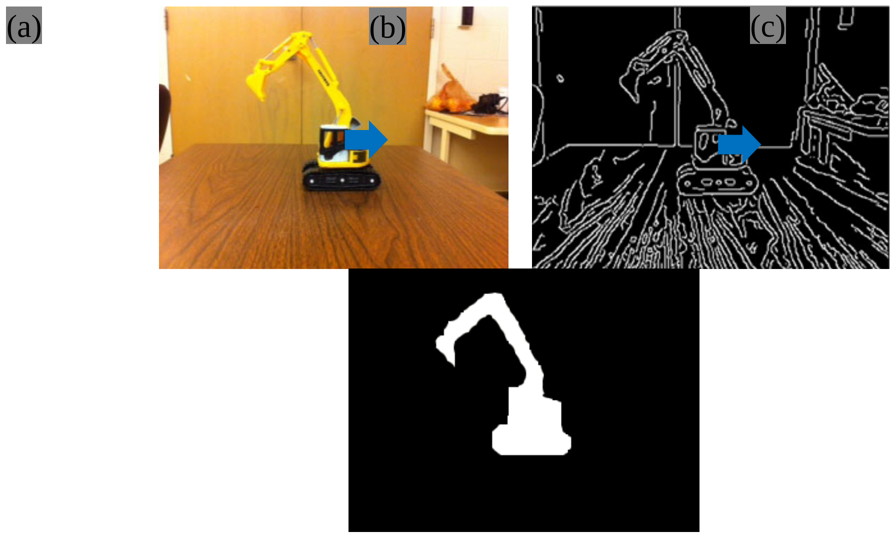


Figure 2. Edge extraction and Segmentation for training data

Key Nodes Modeling and Kinematic Features Extraction. Key nodes modeling aims to generate abstract kinematic features (i.e. nodes) and relations (i.e. edges) among a set of rotatable parts (e.g., booms and bucket). It has great potential to use key nodes model to represent the object's movement. Figure 3 shows the evolution of a tree structure which uses the joint nodes (red dots) and the end nodes (hollow dot) as the key nodes. Considering the visibility frequency, the joint node B is selected as the root node of the tree which should be first detected for the objects detection process. As the bucket is always embedded when digging into soil, it is allowed to add empty branch (edge AD) and endpoint (D) during the process of construction of the tree.

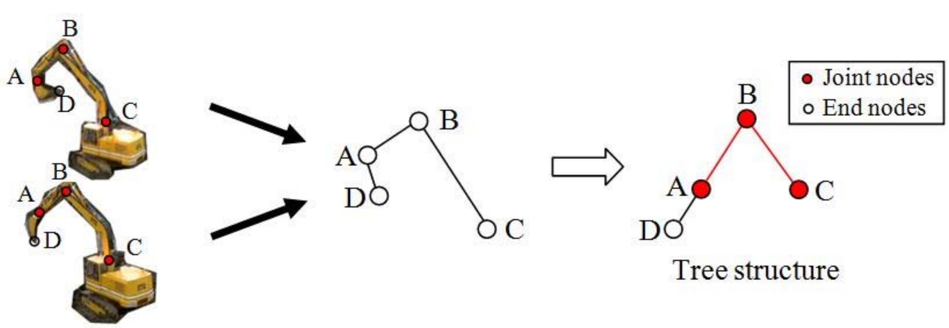
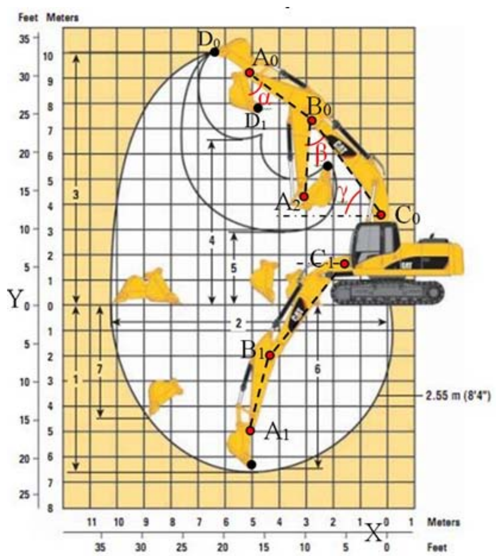


Figure 3. Two excavators in different poses with the same key nodes model

In order to extract accurate kinematic features and incorporate into the key nodes model, the specification of CAT 336D L hydraulic excavator (CAT) was investigated as an example. One of the scale-invariant features for object recognition in 2D images is the angles of the joints which connect the operator station, booms and the bucket. Figure 4 shows the ranges for angles α, β, γ from the spec (CAT), which can be used as the rotation constraints in the key nodes model.



Node	X(meter)	Y(meter)
A_0	5.0	9.2
A_1	5.0	-5.0
A_2	3.0	4.2
B_0	2.8	7.4
B_1	4.4	-2
C_0	0.2	3.5
C_1	1.5	1.5
D_0	6.4	10.0
D_1	4.8	7.8

Angle	Min	Max
α	43°	170°
β	37°	160°
γ	-50°	56°

Figure 4. Boom working ranges for CAT 336D L hydraulic excavator (CAT)

Since it is not an easy task to observe the key nodes from the image, a more intuitive feature — 2D contour/edge, was used as the intermediate templates to derive the tree structure. In the following sub-section, the process of manually producing multiple contour templates and the relationship between contour and key nodes will be discussed.

Contour Templates. In order to manually produce supervised contour templates, first one excavator image is selected and the contour of its bucket and two booms are extracted using Canny detector algorithm (Fig.5, left part). Then each contour component is rotated within the ranges of angles α, β, γ respectively (Fig.5 right), and finally the rotated contour components of the bucket and two booms are combined to produce a new contour template. Meanwhile, the key nodes can be found by either searching the overlaid regions of connected contour components (A, B) or simply searching the end of bucket and the long boom (C, D) (Fig.5 left).

Figure 6 shows a simple training test which extracted the contour feature based on one key node model using the fan shape training procedure proposed by Xinggang et al. (2012). In future study, more key nodes should be concatenated.

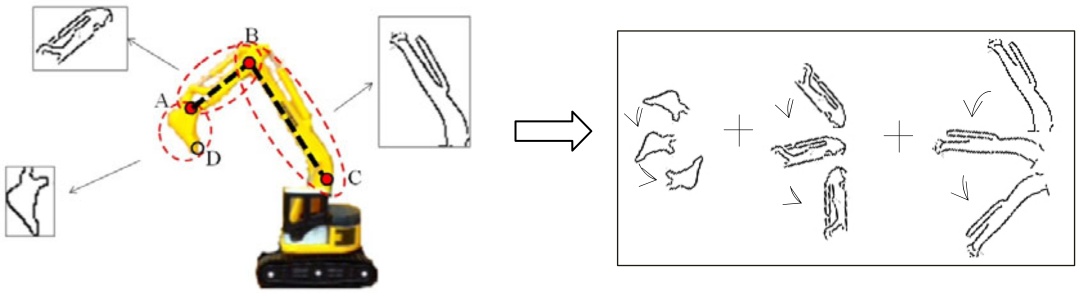


Figure 5. Contour templates design under supervision (Left: relation between contour components and key nodes; Right: different angles of three contour components)

Figure6. A simple training test based on one key node model

Feature Matching. Feature matching process incorporates both merits of key nodes model and contour templates. Figure 7 shows the conceptual workflow of the detection process using sum-and-max algorithm(Bai et al. 2009). An edge image of toy excavator is used to demonstrate this concept. Starting from the root node of the tree (B node), contour components are matched by maximizing the similarity (Sum1) between detected image and contour template components. Maximum responses of the candidate templates connecting to the current node are kept (Max1). Then the process moves to the next node (joint nodes A and C) in the tree to match with other components (Max2, 3). As C is the end node in the right branch, once retrieved, the right process is done. For left branch, after A is searched, the process will continue to the next D node. After all branches are retrieved, these local maximum responses will sum up to an overall matching result (SUM).

Finally, the matched template is determined and therefore the key nodes can be detected and located in the image space.

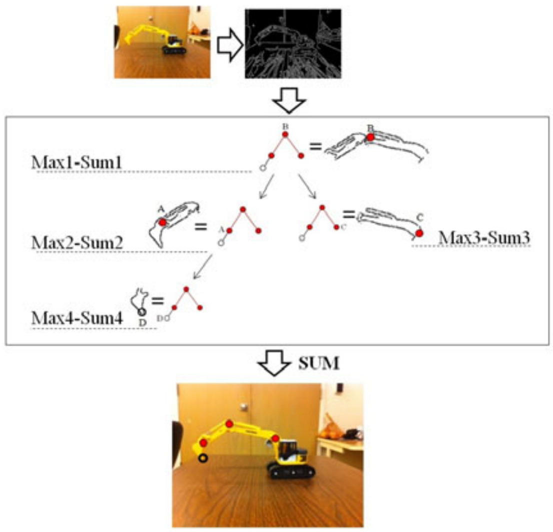


Figure 7. Conceptual workflow of the detection process using sum-and-max algorithm

Key Nodes Location. From the above three steps, we can obtain the key node pixel value in images or video frames. The final task is to determine the 3D coordinates of these key nodes by linking the image pixel value and 3D point cloud depth value. However, as the influence of both image noise and estimation errors, the returned pixel value will be a region instead of a specific point. In this situation, we need to retrieve this region in the corresponding point cloud dataset and remove the outliners in order to find the most confident value as the 3D spatial coordinates for these key nodes.

SUMMARY AND CONCLUSIONS

Object detection and tracking is a challenging problem in the dynamic construction environment, especially for objects with kinematic joints and changing poses, e.g. excavators. This paper presents a kinematic key nodes model for dynamically detecting and locating movable objects in construction sites using color-depth cameras. The key nodes model is designed by first analyzing kinematic constraints from mechanic specifications of interested construction site objects and then building representative segmented contour templates under supervision. Feature matching is realized by comparing the real object components and segmented contour templates in a traversal order of the tree structure for all the key nodes. This knowledge-based key nodes model and feature matching method showed great potential for construction monitoring and spatial safety analysis in an effective and efficient way.

Future work will include both the indoor and outdoor validation to test the feasibility and accuracy of the proposed conceptual methodology.

ACKNOWLEDGEMENTS

The authors thank Shuai Li, Peng Mao and Jason Park for their valuable suggestions during the paper composition. Any opinions, findings and conclusions in this paper are those of the authors and do not necessarily reflect the views of Purdue University.

REFERENCES

Argüelles-Fraga, R., Ordóñez, C., García-Cortés, S., and Roca-Pardiñas, J. (2013). "Measurement planning for circular cross-section tunnels using terrestrial laser scanning." *Automation in Construction*, 31(0), 1-9.

Bai, X., Wang, X., Latecki, L. J., Liu, W., and Tu, Z. "Active skeleton for non-rigid object detection." *Proc., Computer Vision, 2009 IEEE 12th International Conference on*, IEEE, 575-582.

Bosché, F. (2012). "Plane-based registration of construction laser scans with 3D/4D building models." *Advanced Engineering Informatics*, 26(1), 90-102.

Canny, J. (1986). "A computational approach to edge detection." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(6), 679-698.

CAT "CAT 336D L hydraulic excavator." <http://xml.catmms.com/servlet/ImageServlet?imageId=C480968>.

Gai, M., Cho, Y. K., and Xu, Q. "Real-time 3D Visualization of Heavy Construction Equipment Operations Using LADAR." *Proc., Computing in Civil Engineering (2013)*, ASCE, 889-896.

Gai, M., and Xu, Q. "Target-Free Automatic Point Clouds Registration Using 2D Images." *Proc., Computing in Civil Engineering (2013)*, ASCE, 865-872.

Gore, S., Song, L., and Eldin, N. (2011). "Photo-modeling for Construction Site Space Planning." University of Houston.

Hinze, J. W., and Teizer, J. (2011). "Visibility-related fatalities related to construction equipment." *Safety Science*, 49(5), 709-718.

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

MacCollum, D. V. (1995). *Construction safety planning*, John Wiley & Sons.

OpenCV2.4.7.0 "Canny Edge Detector." <http://docs.opencv.org/doc/tutorials/imgproc/imgtrans/canny_detector/canny_detector.html>.

OSHA (2013). "Census of Fatal Occupational Injuries Summary, 2012." <http://www.bls.gov/news.release/cfoi.nr0.htm>. (August 22, 2013).

Rezazadeh Azar, E., and McCabe, B. (2012). "Automated visual recognition of dump trucks in construction videos." *Journal of Computing in Civil Engineering*, 26(6), 769-781.

Tang, P., and Alaswad, F. S. "Sensor Modeling of Laser Scanners for Automated Scan Planning on Construction Jobsites." *Proc., Construction Research Congress 2012*, ASCE, 1021-1031.

Uslu, B., Golparvar-Fard, M., and de la Garza, J. M. "Image-based 3D reconstruction and recognition for enhanced highway condition assessment." *Proc., Proceedings of the 2011 ASCE Intl. Workshop on Computing in Civil Engineering, Miami, FL*, 67-76.

Xinggang, W., Xiang, B., Tianyang, M., Wenyu, L., and Latecki, L. J. "Fan Shape Model for object detection." *Proc., Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 151-158.

Zou, J., and Kim, H. (2007). "Using hue, saturation, and value color space for hydraulic excavator idle time analysis." *Journal of computing in civil engineering*, 21(4), 238-246.