ICCEPM 2020
The 8th International Conference on Construction Engineering and Project Management
Dec. 7-8, 2020, Hong Kong SAR

Image Processing-based Object Recognition Approach for Automatic Operation of Cranes

Ying Zhou[1]*, Hongling Guo[2] , Ling Ma[3], Zhitian Zhang[4]

[1] *Department of Construction Management, Tsinghua University, Beijing,* E-mail address: zhouying17@mails.tsinghua.edu.cn

[2] *Department of Construction Management, Tsinghua University, Beijing,* E-mail address: hlguo@tsinghua.edu.cn

[3] *Department of Construction Management, Tsinghua University, Beijing,* E-mail address: maling901@163.com

[4] *Department of Construction Management, Tsinghua University, Beijing,* E-mail address: zhangzt18@mails.tsinghua.edu.cn

Abstract: The construction industry is suffering from aging workers, frequent accidents, as well as low productivity. With the rapid development of information technologies in recent years, automatic construction, especially automatic cranes, is regarded as a promising solution for the above problems and attracting more and more attention. However, in practice, limited by the complexity and dynamics of construction environment, manual inspection which is time-consuming and error-prone is still the only way to recognize the search object for the operation of crane. To solve this problem, an image-processing-based automated object recognition approach is proposed in this paper, which is a fusion of Convolutional-Neutral-Network (CNN)-based and traditional object detections. The search object is firstly extracted from the background by the trained Faster R-CNN. And then through a series of image processing including Canny, Hough and Endpoints clustering analysis, the vertices of the search object can be determined to locate it in 3D space uniquely. Finally, the features (e.g., centroid coordinate, size, and color) of the search object are extracted for further recognition. The approach presented in this paper was implemented in OpenCV, and the prototype was written in Microsoft Visual C++. This proposed approach shows great potential for the automatic operation of crane. Further researches and more extensive field experiments will follow in the future.

Key words: automated object recognition, vertex-based determining model, image processing, automatic cranes

1. INTRODUCTION

Recent years, with the rapid development of information technology, automatic construction has attracted more and more attention in construction industry. As the most essential components on site with large moving and heavy loads, cranes are involved in all kinds of construction tasks ranging from assembling prefabricated components to hoisting construction materials like rebars and timbers [1]. Therefore, considering efficiency and safety, cranes should be more automated in the future.

Automatic crane is not a new problem in construction industry, previous researches including: 1) the selection of crane type [2] [3] [4], 2) the location identification and optimization of mobile crane [5] [6] [7], 3) lifting path planning [8] [9] [10], 4) coordination of multiple cranes [11] [12] [13], 5) the simulation and visualization for crane operation [14] [15]. We can find that there are few researches about the automatic recognition of crane which is not only the base of automatic crane, but also necessary to achieve fully automatic level, assuming the initial position of the search object for cranes is given [16]. While in practice, limited by the complexity and dynamics of construction environment, conventional manual inspection is the only way on site to determine the original position of the search object, making the operation of crane time-consuming and error-prone.

Recent years, the rapid developments of information technology have provided us with new sensor-based solutions (e.g. three-dimensional (3D) laser scanning and image processing) to realize automated object recognition of crane. And image processing is faster and cheaper with acceptable precision than 3D laser scanning. Besides, 3D laser transmitter is usually loaded on Unmanned Aerial Vehicle (UAV) which however is not allowed to use in some real cases due to the limitation of local regulations, considering government security and privacy protection. Therefore, image processing is selected as the basic automation recognition technology in this research.

2. LITERATURE REVIEW

In order to realize automatic object recognition of crane, its definition should be clarified at first. Although object recognition is a basic problem that has been extensively mentioned in previous research, recognition approaches are different in specific contexts (e.g. computer vision [17] [18] [19], robotics [20] [21] and autonomous driving [22] [23]). In general, object recognition in previous research is defined as the location of a desired object in a scene [24] including: 1) the detection and classification in 2D image, 2) 3D location of the search object with orientation, motion, pose and structure, based on a prior knowledge of the search object including its shape, color and other features.

However, considering the following differences between construction and the applications mentioned above: 1) most recognition targets in construction are artificial objects with similar concrete textures and regular shapes; 2) the 3D location with exact pose of the search object is necessary to determine the following lifting path planning of the crane; 3) except for visible features mentioned above (e.g. texture), other invisible features such as weight [5] are also vital to automatic lifting. Based on this, the object recognition of automatic cranes in this research can be defined as the fusion of object detection in 2D image and the 3D location of the search object with determined visible features (e.g., location, pose, size, and texture) and invisible features (e.g., weight).

Traditionally, feature-based object detection methods are prevalent because of their simplification, speediness and accuracy. Based on the different types of detection template, existing feature-based detection methods can be classified as Histogram-of-Oriented-Gradient (HOG)-based [25], texture/color-based [26] [27], shape-based [28] [29] [30] and scale/affine-invariant feature-based [31] [32]. However, these methods are not robust enough in practice due to the unavoidable effects of the changing environment on feature extraction. In addition, most feature extraction algorithms applied (e.g. Canny operator [33] and Hough detection [34]) are unadjustable image processing in practice with fixed thresholds, making it unsuitable to the complicated construction environment. Although traditional object detection methods in 2D image have shown great potential in feature extraction with speediness and accuracy, they are not robust enough to resist the effects on feature extraction caused by the changing external circumstances, limited to the fixed thresholds.

Recently, the rapid developments of machine learning and deep learning have also promoted the application of Computer Vision (CV), especially Convolutional-Neutral-Network (CNN)-based object detection in construction, as shown in Table 1. According to the research, previous CNN-based detection researches mainly focus on damage/defect detection [30] [35] [36], object classification [37] [40] and workers' activities recognition [38] [39] on site, few exploiting the detection results for further object location in 3D, although it has shown great potential in extracting the search object from complicated 2D backgrounds.

Table 1. CNN-based detection researches in construction

Year	Objective	Applied detection algorithms	Results	Paper
2010	Concrete columns detection for automated bridge inspection	Artificial neutral network for concrete material recognition, Canny operator, Hough transform	Accuracy of 89.7%	[30]
2015	Changes detection in tunnel lining	CNN	-	[35]

2018	Detecting healthy surface regions of bridges	Sliding window approach, GoogleNet Inception v3 network	Accuracy of 82.8%, a search space reduction of 90.1%	[36]
2018	No-hardhat-use detection	Faster R-CNN	Precision more than 90%	[37]
2018	Recognizing workers' activities	Two-stream convolutional networks	An average accuracy of 80.5%	[38]
2018	Workforce activity assessment for steel reinforcement fixers	An improved CNN	An average accuracy of 85%	[39]
2018.5	Detect the presence of workers and excavators	Improved Faster R-CNN	As a high level of accuracy (91% and 95% respectively)	[40]

In conclusion, no matter feature/non-feature-based or CNN-based, all object detection approaches mentioned above only classify and detect objects of interest in 2D images, which means that spatial location of the search objects with exact pose is little considered in previous researches. Hence, a novel and robust object recognition approach is proposed in this paper, which exploits the synchronized relationship of 3D spatial location and the 2D detection results of CNN-based and traditional object detection from sensed real-time data to locate the search object with exact pose, more accurate visible features and invisible features. The primary objective of the approach presented in this research is to examine the feasibility of image-processing-based object recognition approach for automatic crane.

3. RESEARCH METHOD

According to the above definition of object recognition for construction cranes and the review of previous research, the following two issues need to be considered seriously so as to develop an appropriate object recognition approach.

3.1. The effects of background and the occlusions of other objects

On-site construction environment is always chaotic with not only the search objects, but also the unsearch objects and other moving workers that may produce occlusions and reduce the efficiency of object recognition. Therefore, how to eliminate the effects of backgrounds and the occlusions of other objects is the first fundamental problem to solve. In this research, considering that 5 fps speed of Faster R-CNN can almost meet the speed requirements of on-site object recognition with acceptable accuracy, Faster R-CNN is selected to realize the basic 2D detection of the search object from complicated backgrounds for further image processing.

3.2. Unknown pose and quantity of the search objects

Object recognition systems genuinely assume that the pose and number of the search objects are a prior unknown knowledge, and the computational complexity used to be proportional to the total quantity of the search objects in traditional recognition methods [21]. Therefore, Faster R-CNN is used to detect the search object individually each with single bounding box, and then a circular image process including simple and efficient Vertex-based determining models is performed to spatially locate the search object with pose.

According to the basic problems proposed above, our approach can be divided into three parts, and Figure 1 depicts the whole flowchart:

1) Real-time information collection of the search object by the chosen sensor (e.g. stereo camera) including synchronized 3D coordinates, image pixels and color information. And then for the retrieved 2D image, Faster R-CNN is selected to segment each object of interest from complicated background;

2) Circular image processing including Canny detection, Hough transformation, endpoints clustering analysis and Vertex-based determining model is performed until the extracted results satisfy the Vertex-based determining model;

3) Features extraction including centroid coordinate calculation, size calculation and color extraction is based on the range image fusing the retrieved 2D image and 3D coordinates of point clouds.

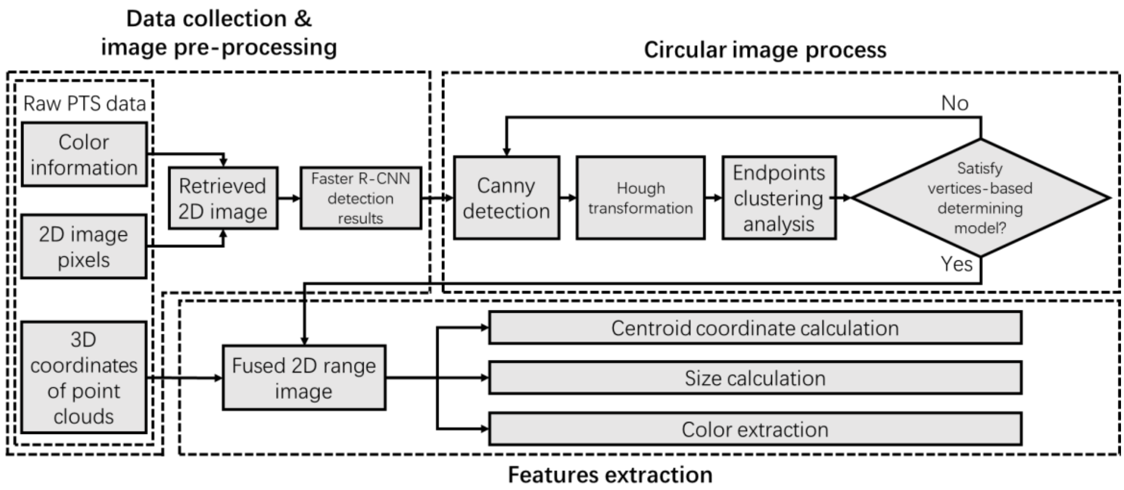
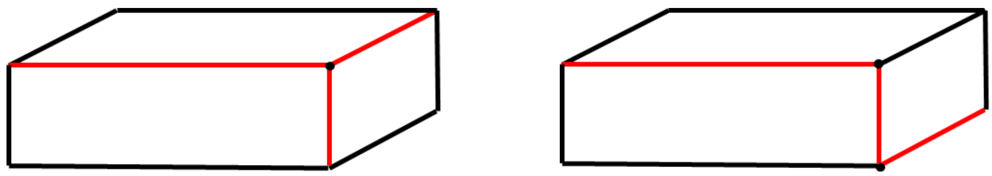


Figure 1. Overview flowchart of the proposed recognition approach

4. OBJECT RECOGNITION APPROACH

To describe the approach used clearly, cuboid-shape lifting objects (such as precast columns and slabs), which are common on construction sites, are used as the experimental prototype. Object recognition of automatic crane in this research is the combination of 2D detection and 3D location which however is based on the image detection results. According to the theory of solid geometry, the spatial location of a cuboid with exact pose can be determined, assuming that at least three specified mutually perpendicularly intersecting line segments are given. These specified lines at the same time intersect at one vertex (see Figure 2(a)) or two vertices (see Figure 2(b)). In another word, as long as the image processing results, the extracted vertex groups, satisfy our proposed Vertex-based determining models, the search object can be uniquely located in 3D space.



(a) The specified lines intersect at one vertex

(b) The specified lines intersect at two vertices

Figure 2. Specified mutually perpendicularly intersecting line segments with vertices to uniquely locate a cuboid in 3D space

4.1. Data collection and image preprocessing

Bumblebee series camera is a multi-view stereo vision component for fast stereo reconstruction developed by FLIR System Inc. Considering its accuracy and stability, it is selected in this research to collect real-time raw depth data of the scene, saved as PTS files where (x, y, z) represent 3D spatial coordinates, (u, v) represent 2D image pixels and (R, G, B) represent pixel color information. Because original images generated by single left or right shots don't contain necessary 3D information (e.g. 3D spatial coordinates), 2D images should be firstly retrieved from the collected raw depth data (u, v, R, G, B) so that the retrieved image is matched with 3D spatial coordinates (x, y, z), as shown in Figure 3(a).

The backgrounds of the search object in the retrieved 2D image are usually complicated because of great turbulences generated by piles of materials, running machines and moving workers. As the computational complexity of image processing and feature extraction is proportional to the interference of background, the search object should be segmented from background to decrease the image size to process. Considering this problem, CNN-based object detection is applied in our research to detect the search objects from complex backgrounds before image processing. And Faster R-CNN is finally selected because of its high processing speed and detection accuracy. And the Foreground detection result of the trained Faster R-CNN is shown in Figure 3(b), with a learning rate of 0.0003, accuracy of the model finally reaching 93.3%.

(a) Retrieved 2D image (b) Foreground detection result of Faster R-CNN

Figure 3. 2D image retrieved from PTS file and foreground detection result of Faster R-CNN

4.2. Image processing

Figure 4 depicts the circular image processing of which the inputs are the foreground detection results of Faster R-CNN. The fragments of the detected bounding boxes are then orderly processed by Canny detection, Hough transformation and Endpoints clustering analysis, to obtain edge contours, line segments with endpoints and vertex groups. As long as the extracted vertex groups satisfy any of the proposed vertex-based determining models, the search cuboid can be uniquely located in 3D. Otherwise it will go back to Canny Detection with adjusted threshold until the extracted vertex groups satisfy the presented models.

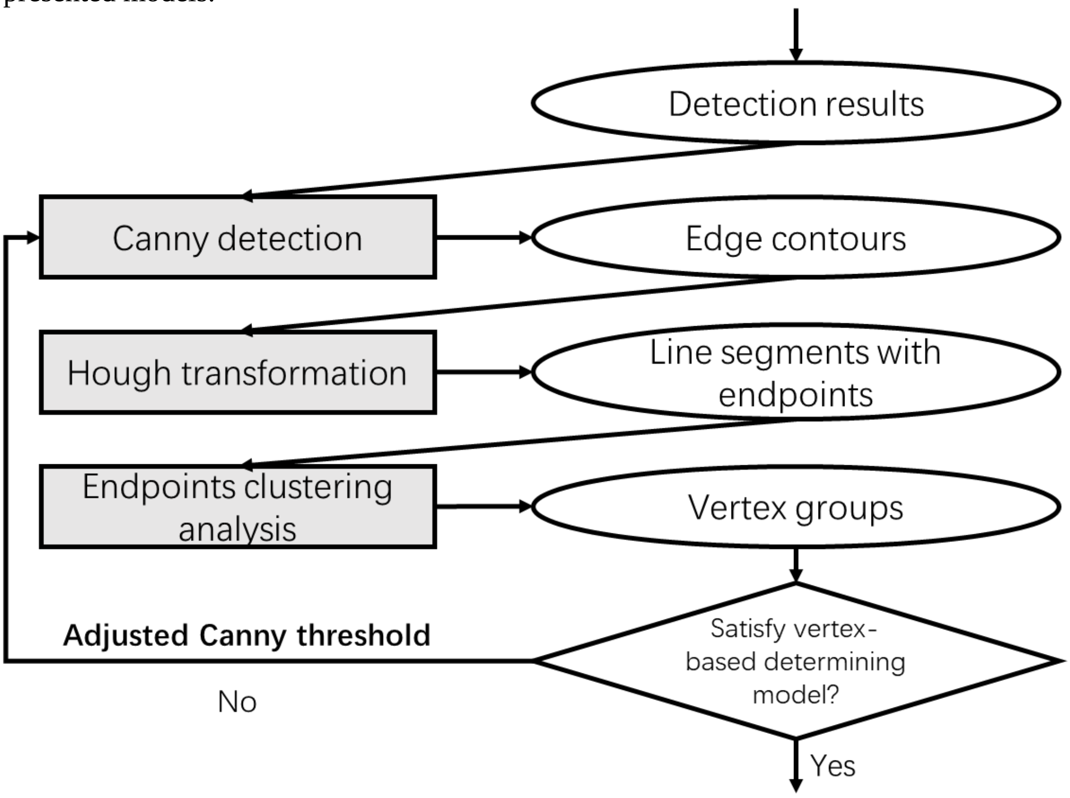


Figure 4. Circular image processing

4.2.1. Canny detection

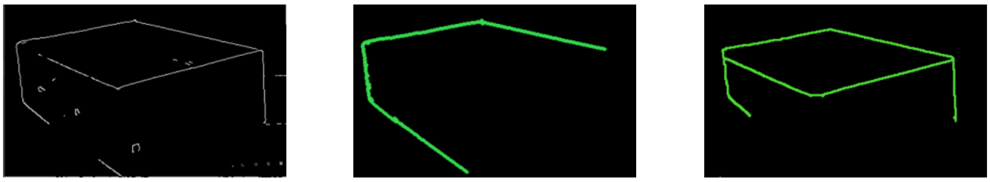
Canny edge detection operator was proposed by Canny [41] in 1986 to obtain the extreme points of image gradient, which are possible edge contours. Based on the obtained extreme points, the true edge contours of the search objects are detected more accurately by non-maximum suppression and double threshold detection. However, two obvious limitations exist in original Canny algorithm——

1) The surface textures of the search objects left in bounding box affect the edge detection results;

2) Fixed Canny threshold for different recognition objects in all circumstances shows weak robustness.

In order to solve the problems above, clustering points which probably are textures instead of edges are detected and eliminated. As shown in Figure 5, the improved Canny with automatically adjusted threshold shows high potential in eliminating the texture contours of the search objects to obtain more accurate edge contours of the search objects. And then a circular image processing from Canny

Detection, Hough Transformation, Endpoints clustering analysis to Vertex-based determining model is presented, with adjusted Canny threshold according to the processing result until the processed vertex groups satisfy the determining models.



(a) Original Canny detection result
(b) Improved Canny detection result I
(c) Improved Canny detection result II

Figure 5. The comparison of edge contours detection results between Canny and improved Canny

4.2.2. Hough transformation

It should be noted that the edge contours extracted by Canny above actually are a series of contour points with extremely close distances, so these detected edge contours are not real line segments. Considering this, by transforming the detection problem of the given curve in the original image into the search of peak point in the special parameter space, Hough transform is adopted to extract the line segments information contained in these contour points. Therefore, through Hough transformation, the edge contours extracted by Canny can be transformed into corresponding line segments and corresponding endpoints with 2D image pixels (u,v). And based on the synchronized relationship between 3D spatial coordinates and 2D image pixels in PTS files, the 3D spatial coordinates (x, y, z) of these endpoints can also be further determined. Table 2 presents the detailed 3D information of a line segment extracted by Hough transformation.

Table 2. The detailed 3D information of the extracted line segments

		x	**y**	**z**	**u**	**v**
Line i	Point 1	$x_{i,1}$	$y_{i,1}$	$z_{i,1}$	$u_{i,1}$	$v_{i,1}$
	Point 2	$x_{i,2}$	$y_{i,2}$	$z_{i,2}$	$u_{i,2}$	$v_{i,2}$
	Spatial direction vector (Δ)	$\Delta x_i = x_{i,2} - x_{i,1}$	$\Delta y_i = y_{i,2} - y_{i,1}$	$\Delta z_i = z_{i,2} - z_{i,1}$	-	-
	Length of Euclidean distance (E)		$E_i = \sqrt[2]{\Delta x_i^2 + \Delta y_i^2 + \Delta z_i^2}$		-	-

The ideal transformation result of Hough is that the extracted line segments with endpoints are the exact edges with vertices of the search object. In this case, adjacent edges of the search object intersect at the vertices, which means that the extracted adjacent line segments should also intersect at their endpoints. However, effected by all kinds of image noise, these extracted line segments in fact do not coincide exactly with the true edges of the search object, with a few deviations between the endpoints and the actual vertices. Therefore, the extracted adjacent line segments do not exactly intersect at their endpoints. To solve this problem, endpoints clustering analysis is performed next to identify the possible intersection relation of these line segments, by merging the endpoints of which the distances is within the pre-set distance threshold.

4.2.3. Endpoints clustering analysis

Based on the extracted line segments and endpoints of Hough Transformation, clustering analysis is performed to calculate the distances between all endpoints of two line segments, by comparing the calculated distances with the pre-set distance threshold D. If the distance is less than or equal to D, the two endpoints are regarded as a clustering point, where two corresponding line segments may intersect. In another word, the clustering point is probably a vertex of the search object. Otherwise, it can be inferred that the corresponding line segments do not intersect with each other. As shown in Figure 6, for the endpoints clustering analysis of two line segments each containing two endpoints individually (e.g. Line i and Line j), there are four groups of endpoints relationships in total to analyze:

1) the relationship between Line i-Point 1 and Line j-Point 1;

2) the relationship between Line i-Point 1 and Line j-Point 2;
3) the relationship between Line i-Point 2 and Line j-Point 1;
4) the relationship between Line i-Point 2 and Line j-Point 2.

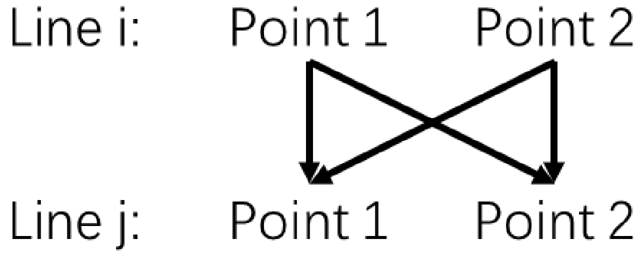


Figure 6. The endpoints clustering analysis of two line segments

The clustering point judgment of two endpoints is based on their Euclidean distance. For example, E_(i1,j2) represents the Euclidean distance between Line i-Point 1（xi,1, yi,1, zi,1）and Line j-Point 2（xj,2, yj,2, zj,2）:

$$E_{i1,j2} = \sqrt{(x_{i,1} - x_{j,2})^2 + (y_{i,1} - y_{j,2})^2 + (z_{i,1} - z_{j,2})^2} \tag{1}$$

If $R_{i1,j2}$ is less than the pre-set distance threshold D (D=0.03 m in this research, which means that endpoints with Euclidean distance ≤ 0.03 m can be regarded as a clustering point), Line i-Point 1 and Line j-Point 2 are a group of clustering point and their information will be recorded. Otherwise, Line i-Point 1 and Line j-Point 2 are not clustered, and the next group of endpoints will be calculated and judged. In this way, clustering point judgment of Line i and Line j is performed on four groups of endpoints relationships in turn. As long as any calculated European distance of the four groups is less than D, it can be inferred that Line i and Line j intersect at the corresponding clustering point (the same vertex group). Otherwise, Line i and Line j don't intersect.

4.2.4. Vertex-based determining model

According to the theory of solid geometry, at least three mutually perpendicular intersecting line segments are needed to uniquely determine the spatial location of the search object, including the following two situations:

1) Three intersecting line segments are perpendicular to each other (see Figure 2(a));
2) Three line segments intersecting at two different vertices are perpendicular to each other (see Figure 2(b)).

Combining with the vertex groups generated by endpoints clustering analysis, the following two vertex-based determining models are proposed.

4.2.4.1 Model Ⅰ

For the first situation, the search problem of three intersecting line segments can be converted to the search of a vertex group at which more than three line segments intersect, that is to say, the aggregation degree of the selected vertex group ⩾3. As for the judgment problem of these line segments intersecting vertically, it can be further converted to determining whether the three line segments containing the selected vertex are perpendicular to each other. Besides, spatial direction vectors Δ of the line segments (see Table 1) can be used for determining whether two line segments (e.g. Line i and Line j) intersect vertically. For example, if

$$\Delta x_i * \Delta x_j + \Delta y_i * \Delta y_j + \Delta z_i * \Delta z_j = 0 \tag{2}$$

Considering the experimental error and measurement error, this condition can be relaxed to

$$|\Delta x_i * \Delta x_j + \Delta y_i * \Delta y_j + \Delta z_i * \Delta z_j| \le 0.05 \tag{3}$$

It can be inferred that Line i and Line j are perpendicular to each other. Otherwise, Line i and Line j are not perpendicular.

4.2.4.2 Model Ⅱ

As for the second situation, the search problem of three line segments intersecting at two different vertices can be converted to the search of a line segment with both endpoints substituted with clustered vertex groups. And the judgment problem of three line segments intersecting vertically can also be further converted to calculating the spatial direction vectors Δ of the selected line segments (see Formula 3).

4.3. Feature extraction

Through image processing, three specified mutually perpendicularly intersecting line segments are extracted as edges which can help to uniquely locate the search object in 3D space. And apart from location, other features including size and color are also necessary for the recognition of the search object. Based on the extracted three specified line segments above, the centroid coordinate of the search object, as the original position for the object to hoist, can be calculated according to triangular rule, and the sizes (length, width and height) are also determined at the same time. As for color, the effects of light on color extraction are considered and eliminated by white balance with satisfactory results.

5. CONCLUSION

An image-processing-based automated object recognition approach for automatic crane is presented in this research. The search object is firstly detected with bounding box from the complicated backgrounds on the retrieved 2D images by Faster R-CNN, which is trained and validated by a collected dataset with accuracy of 93.3%. Based on the detection results, a circular image processing including different object processing technologies (Canny detection, Hough transformation and Endpoints clustering analysis) is performed to extract all vertex groups of the search object. If the extracted vertex groups satisfy any of the two proposed Vertex-based determining models, the search cuboid is uniquely located in 3D with exact pose. Otherwise it will go back to Canny detection with adjusted threshold, until the extracted vertex groups satisfy the proposed models. Then three vital visible features of the search object including centroid coordinate, size and color can be further extracted.

However, there are still some limitations need to be improved in future research:

1) Decrease the effects of light on texture extraction. Due to the differences in light sources, weathers and shooting times, texture/color of the same search objects in different circumstances may be quite different from each other, with the accuracy of color extraction results decreased.

2) The address of invisible attributes. Except for visible features (e.g. color and size), other invisible attributes of the search object (e.g. weight) are also necessary for the automatic lifting of crane, which however are stored in IFC scheme of BIM models and unavailable directly by image processing technologies.

Although this proposed approach shows great potential for the automatic operation of crane, further researches and more extensive field experiments need to follow in the future.

ACKNOWLEGEMENTS

We would like to thank the National Natural Science Foundation of China (Grant No. 51578318, 51208282) as well as Tsinghua University-Glodon Joint Research Centre for Building Information Model (RCBIM) for supporting this research. Besides, thanks for the help provided by Zhubang Luo in computer programming.

REFERENCES

[1] Z. Lei, M. Al-Hussein, U. Hermann, A. Bouferguene, "Heavy lift analysis at FEED stage for industrial project", 2016 Winter Simulation Conference (WSC), pp. 3281-3289, 2016.

[2] C. Gray, J. Little, "A systematic approach to the selection of an appropriate crane for a construction site", Construction Management and Economics, vol. 3, no.2, pp. 121-144, 1985.

[3] C. W. Farrell, K. C. Hover, "Computerized crane selection and placement for the construction site", The 4th International Conference on Civil and Structural Engineering Computing, vol. 1, pp. 91-94, 1989.

[4] Z. Pan, H. Guo, Y. Li, "Automated Method for Optimizing Feasible Locations of Mobile Cranes Based on 3D Visualization. Procedia engineering", Procedia engineering, vol. 196, pp. 36-44, 2017.

[5] U. Hermann, A. Hendi, J. Olearczyk, M. Al-Hussein, "An integrated system to select, position, and simulate mobile cranes for complex industrial projects", Construction Research Congress 2010, pp. 267-276, 2010.

[6] H. Safouhi, M. Mouattamid, U. Hermann, A. Hendi, "An algorithm for the calculation of feasible mobile crane position areas", Automation in Construction, vol. 20, no. 4, pp. 360-367, 2011.

[7] M. A. D. Ali, N. R. Babu, K. Varghese, "Collision free path planning of cooperative crane manipulators using genetic algorithm", Journal of computing in civil engineering, vol. 19, no. 2, pp. 182-193, 2005.

[8] Z. Lei, S. Han, A. Bouferguène, H. Taghaddos, U. Hermann, M. Al-Hussein, "Algorithm for mobile crane walking path planning in congested industrial plants", Journal of Construction Engineering and Management, vol. 141, no. 2, pp. 05014016, 2015.

[9] Y. Lin, H. Yu, G. Sun, P. Shi, "Lift Path Planning without Prior Picking/Placing Configurations: Using Crane Location Regions", Journal of Computing in Civil Engineering, vol. 30, no. 1, pp. 04014109, 2016.

[10] S. C. Kang, E. Miranda, "Computational methods for coordinating multiple construction cranes", Journal of Computing in Civil Engineering, vol. 22, no. 4, pp. 252-263, 2008.

[11] C. Zhang, A. Hammad, "Multiagent approach for real-time collision avoidance and path replanning for cranes", Journal of Computing in Civil Engineering, vol. 26, no. 6, pp. 782-794, 2012.

[12] W. Ren, Z. Wu, "Real-time anticollision system for mobile cranes during lift operations", Journal of Computing in Civil Engineering, vol. 29, no. 6, pp. 04014100, 2015.

[13] S. Han, A. Bouferguene, M. Al-Hussein, U. Hermann, "3D-based crane evaluation system for mobile crane operation selection on modular-based heavy construction sites", Journal of Construction Engineering and Management, vol. 143, no. 9, pp. 04017060, 2017.

[14] S. H. Han, S. Hasan, A. Bouferguène, M. Al-Hussein, J. Kosa, "Utilization of 3D visualization of mobile crane operations for modular construction on-site assembly", Journal of Management in Engineering, vol. 31, no. 5, pp. 04014080, 2015.

[15] K. He, G. Gkioxari, P. Dollár, R. Girshick, "Mask r-cnn", 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980-2988, 2017.

[16] S. Kang, E. Miranda, "Planning and visualization for automated robotic crane erection processes in construction", Automation in Construction, vol. 15, no. 4, pp. 398-414, 2006.

[17] R. Girshick, "Fast r-cnn", 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448, 2015.

[18] S. Ren, K. He, R. Girshick, J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137 - 1149, 2017.

[19] T. S. Huang, A. N. Netravali, "Motion and structure from feature correspondences: A review", Proceedings of the IEEE, vol. 82, no. 2, pp. 252 - 268, 1994.

[20] F. Bosche, C. T. Haas, "Automated retrieval of 3D CAD model objects in construction range images", Automation in Construction, vol. 17, no. 4, pp. 499-512, 2008.

[21] A. Geiger, P. Lenz, R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite", 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354-3361, 2012.

[22] G. N. DeSouza, A. C. Kak, "Vision for mobile robot navigation: A survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 2, pp. 237-267, 2002.

[23] F. Ge, T. Liu, S. Wang, J. Stahl, "Template-based object detection through partial shape matching and boundary verification", International Journal of Electrical and Computer Engineering, vol. 2, no. 11, pp. 2562-2571, 2008.

[24] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005.

[25] J. Abeid Neto, D. Arditi, M. W. Evens, "Using Colors to Detect Structural Components in Digital Pictures", Computer - Aided Civil and Infrastructure Engineering, vol. 17, no. 1, pp. 61-67, 2002.

[26] I. K. Brilakis, L. Soibelman, Y. Shinagawa, "Construction site image retrieval based on material cluster recognition", Advanced Engineering Informatics, vol. 20, no. 4, pp. 443-452, 2006.

[27] Z. Zhu, I. Brilakis, "Detecting air pockets for architectural concrete quality assessment using visual sensing", Electronic Journal of Information Technology in Construction, vol. 13, pp. 86-102, 2008.

[28] Z. Zhu, I. Brilakis, "Concrete Column Recognition in Images and Videos", Journal of Computing in Civil Engineering, vol. 24, no. 6, pp. 478-487, 2010.

[29] Z. Zhu, S. German, I. Brilakis, "Detection of large-scale concrete columns for automated bridge inspection", Automation in Construction, vol. 19, no. 8, pp. 1047-1055, 2010.

[30] K. Mikolajczyk, C. Schmid, "A performance evaluation of local descriptors", Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03), 2003.

[31] W. Guo, L. Soibelman, J. H. Garrett Jr, "Automated defect detection for sewer pipeline inspection and condition assessment", Automation in Construction, vol. 18, no. 5, pp. 587-596, 2009.

[32] A. Bandera, J. M. Pérez-Lorenzo, J. P. Bandera, F. Sandoval, "Mean shift based clustering of Hough domain for fast line segment detection", Pattern Recognition Letters, vol. 27, no. 6, pp. 578-586, 2006.

[33] A. E. Johnson, M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 5, pp. 433-449, 1999.

[34] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[35] P. Hüthwohl, I. Brilakis, "Detecting healthy concrete surfaces", Advanced Engineering Informatics, vol. 37, pp. 150-162, 2018.

[36] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, W. An, "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos", Automation in Construction, vol. 85, pp. 1-9, 2018.

[37] X. Luo, H. Li, D. Cao, Y. Yu, X. Yang, T. Huang, "Towards efficient and objective work sampling: Recognizing workers' activities in site surveillance videos with two-stream convolutional networks", Automation in Construction, vol. 94, pp. 360-370, 2018.

[38] H. Luo, C. Xiong, W. Fang, P. E. Love, B. Zhang, X. Ouyang, "Convolutional neural networks: Computer vision-based workforce activity assessment in construction", Automation in Construction, vol. 94, pp. 282-289, 2018.

[39] W. Fang, L. Ding, B. Zhong, P. E. Love, H. Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach", Advanced Engineering Informatics, vol. 37, pp. 139-149, 2018.

[40] M. R. Jahanshahi, F. Jazizadeh, S. F. Masri, B. Becerik-Gerber, "Unsupervised Approach for Autonomous Pavement-Defect Detection and Quantification Using an Inexpensive Depth Sensor", Journal of Computing in Civil Engineering, vol. 27, no. 6, pp. 743-754, 2013.

[41] C. John, "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679-698, 1987.