2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Calibration-Free Vision-Assisted Container Loading of RTG Cranes

Jianbing Yang, Yuanzhe Wang, Hao Jiang, Bin Zhao, Yiming Li, Danwei Wang *Fellow, IEEE*

***Abstract*— Vision-assisted container loading of Rubber Tyred Gantry (RTG) cranes are facing two primary challenges. Firstly, the uncertainty inherent in Covolutional Neural Network (CNN) based detection hinders its direct application in the safety-critical operation of such heavy-duty machinery. Secondly, sensor calibration introduces additional complexities and errors into the system. However, existing studies have not adequately addressed these challenges. Motivated by this gap, this paper proposes an integrated approach for target detection and alignment control in container loading of RTG cranes. To ensure reliable target marker identification, a heuristic post-processing algorithm is developed as a complement to CNN-based foreground segmentation, thereby ensuring safety during the container handling process. On this basis, a pixel-based control scheme is designed to align the container with the target markers, which eliminates the need for offline or online sensor calibrations. The proposed approach has been successfully implemented on a real RTG crane manufactured by Shanghai Zhenhua Heavy Industries Co., Ltd. (ZPMC) and validated at the Port of Ningbo, China. Experimental results demonstrate the superiority of the proposed approach over current manual operations in port industries, highlighting its potential for crane automation.**

I. Introduction

Rubber Tyred Gantry (RTG) cranes play an important role in container handling within port operations. Typically, an RTG crane comprises a gantry, a trolley, and a spreader. In practice, container loading operations are initiated by commands from the terminal operating system (TOS) [1], which specify target information for the containers [2]. Upon receiving the commands, the control system coordinates simultaneous movements of the gantry and trolley to position the container roughly over its designated location. Subsequently, the hoist lowers the container into place. However, in practical scenarios, achieving precise container placement solely through gantry and trolley control proves challenging due to variations in air pressure and tire abrasion among the rubber tires. To address this issue, most RTG spreaders are equipped with micro-motion mechanisms to fine-tune their position and attitude for accurate alignment between container and target markers. This paper delves into the container loading problem of RTG cranes.

In the existing literature, works related to RTG crane container loading can be broadly categorized into two main groups: remote operation and automated alignment. Remote operation involves human operators manipulating the micro-motion mechanism based on images captured by onboard cameras of the spreader (as depicted in Fig. 1) to align containers with target markers [3]. However, the effectiveness of this approach heavily relies on the operator's experience, leading to potential variability in control performance. To enhance reliability, automated alignment methods have been introduced, primarily utilizing two types of sensing technologies: *LiDAR-assisted* and *vision-assisted*. LiDAR-assisted alignment is resilient to illumination variations but suffers from drawbacks such as high cost and lack of semantic information. In addition, as a precise instrument, limited mounting positions are suitable for a LiDAR as we need to protect it from high impact and severe vibrations [4]. Conversely, vision-assisted alignment leverages richer semantic information (e.g., shape, color, texture), enabling better target marker detection at a lower cost, making it preferable in the port industry. Nonetheless, vision-assisted container loading still encounters two significant challenges. Firstly, cameras are susceptible to changes in illumination, leading to uncertainties in target marker detection that could compromise crane safety. Secondly, the transformation from image coordinates to world coordinates necessitates intrinsic [5] and extrinsic [6] calibrations, where errors in calibration may propagate to the automated control system. Furthermore, mechanical deformations can substantially alter calibration parameters, exacerbating errors in detection results.

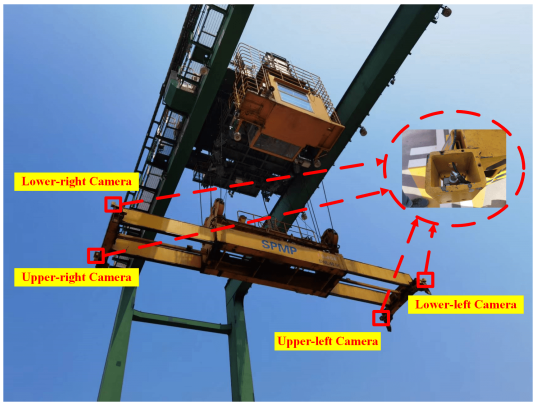


Fig. 1. The cameras mounted on the RTG crane spreader

This research is partly supported by the National Research Foundation, Singapore and the Maritime and Port Authority of Singapore under the Maritime Transformation Programme (Project No. SMI-2022-MTP-04).

J. Yang is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798, and was with the Shanghai Zhenhua Heavy Industries Co., Ltd. (ZPMC), Shanghai, China, 200125.

Y. Wang is with the School of Control Science and Engineering, Shandong University, Jinan, China, 250061, and was with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798.

D. Wang is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798.

H. Jiang, B. Zhao, and Y. Li are with the Shanghai Zhenhua Heavy Industries Co., Ltd., Shanghai, China, 200125.

Corresponding author: Yuanzhe Wang, wang0951@e.ntu.edu.sg

Motivated by the challenges outlined above, this paper proposes a novel integrated approach for target detection and alignment control in vision-assisted container loading of RTG cranes. Comprising two key modules, namely target marker detection and automated alignment control, the proposed methodology aims to overcome the identified challenges effectively. To address the first challenge, a CNN-based target detection algorithm is developed to locate target markers on the ground, supplemented by a heuristic post-processing algorithm to mitigate uncertainties inherent in CNN-based detection. To address the second challenge, a pixel-based alignment control scheme is devised to drive the micro-motion mechanism, ensuring precise alignment of the container with the target markers. The efficacy of the proposed approach is validated through implementation on a real RTG crane manufactured by Shanghai Zhenhua Heavy Industries Co., Ltd. (ZPMC) at the Port of Ningbo, China, with experimental results showcasing its superiority over current manual operations in port industries. The contribution of this paper is the development of an integrated target detection and alignment control approach and the successful implementation on a real RTG crane, distinguishing it from existing works that separately address target detection and alignment control. Furthermore, the proposed approach offers two novel features:

- Integration of deep learning with heuristic image post-processing not only achieves precise detection in outdoor environments but also ensures the absolute correctness of detection results, thereby establishing the system to the high safety standard in port industries and successfully incorporating deep learning into industrial automation control for RTG cranes.
- Elimination of intrinsic and extrinsic calibrations for the spreader onboard cameras simplifies the system and reduces sources of error, enhancing overall efficiency and reliability.

The remainder of this paper is organized as follows. Sec. II reviews related works. Sec. III introduces the proposed approach. Experimental results are presented in Sec. IV, while conclusions and future work recommendations are provided in Sec. V.

II. Related Works

Container loading of RTG cranes consists of two essential steps. Firstly, the landing target need to be identified and localized. Subsequently, the automated control system aligns the container with and loads it to the target. Currently, most existing studies in this domain focus on target detection and have not yet integrated automated control. Based on the employed sensing technology, existing works can be categorized into two main types: *LiDAR-assisted* and *vision-assisted*. LiDAR-assisted container loading utilizes LiDAR as the primary sensor for target detection. References [7], [8] propose a truck positioning system for container loading and unloading onto prime movers, where LiDARs are installed on both sides of the crane's transversal girder. This mounting point vibrates slightly in practice, ensuring safety during long term operations. A drawback of this system is that due to the long distance between the LiDAR and the target, even a slight deformation of the crane girder would cause significant measurement errors. References [9], [10] introduce an auto-landing system developed for container stacking, with LiDARs mounted on the trolley to monitor the relative pose relationship between the spreader and the landing target. However, similar challenges arise due to the substantial distance between the LiDAR and the target, as well as mechanical vibrations and deformations, potentially leading to nonnegligible errors. In practice, cranes may be subject to high impact and severe vibrations. As a precise instrument, the mounting position of LiDAR must avoid locations with such disturbances, thereby limiting the flexibility of sensor deployment. Furthermore, LiDAR lacks semantic information, weakening its capability in target recognition.

Vision-assisted container loading employs cameras to detect landing targets. References [11], [12] utilize trolley onboard CCD cameras to detect specifically designed landmarks on the upper surface of the spreader. The known distance between these landmarks facilitates the estimation of the relative pose relationship between the spreader and the trolley, ensuring accurate spreader localization. References [13], [14] mount two CCD cameras at two diagonal corners of the spreader to capture images of the container's corners. Geometric analysis is then carried out to estimate the relative pose between the spreader and the container. Additionally, a laser mounted at the spreader's center measures the distance between the spreader and the container, enabling the transformation of the relative pose from image coordinates to world coordinates. Reference [15] adopts a pair of stereo cameras on the spreader and employs correlation-based stereo matching techniques to estimate the depth information for the transformation from image coordinates to world coordinates. Vision-based techniques, with their rich semantic information, offer the potential for more accurate target detection. However, most existing studies rely on camera extrinsic [5] and intrinsic calibration [6] technologies for the transformation from image coordinates to world coordinates. In practice, even minor mechanical deformations in the container crane may lead to significant errors of the calibration parameters, resulting in detection failure.

With the rapid advancement of deep learning, numerous object detection and segmentation algorithms have emerged, capable of accurately locating targets in complex environments, despite persistent uncertainties and the black-box characteristics inherent in neural networks. This progress heralds potential opportunities for the application of deep learning based computer vision in outdoor industrial environments. Presently, several straightforward vision-based technologies, including container number recognition [16], damage inspection [17], and detection of hazardous materials symbols [18], have already found implementation in port operations. However, despite these advancements, there is a notable absence of existing research on the application of deep learning based vision technology to the container loading process of container cranes, a safety-critical operation of

heavy-duty machinery.

III. Approach

In manual operations, human operators rely solely on image feedback from the onboard cameras of the spreader to guide the container to the center of the target markers. Throughout this process, operators lack precise information regarding the relative pose of the container with respect to the target markers in world coordinates. Decisions are made based solely on observed motion trends in the image feedback. Inspired by manual operations, we propose an integrated approach for target detection and alignment control to tackle the container loading challenge (as shown in Fig. 2). The target detection module begins with foreground segmentation to extract the target markers, followed by a postprocessing step to determine the precise positions of the marker corner points. Meanwhile, the alignment control module estimates the relative pose relationship between the container and the target markers in image coordinates, generating control commands for the micro-motion mechanism to adjust the pose of the spreader. In the following, the proposed approach will be elaborated.

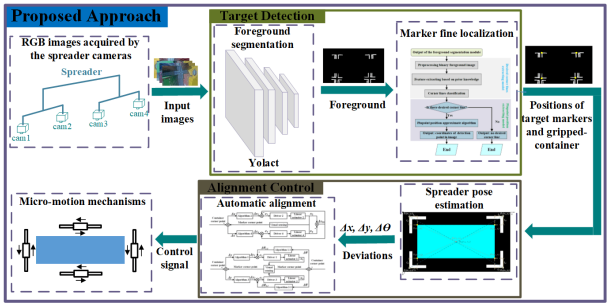


Fig. 2. The diagram of the proposed approach

A. Target Detection

The target detection module extracts the foreground of the target markers from the image and localizes their corner points in image coordinates. Firstly, the target marker is labeled through segmentation annotation, leveraging a dataset comprising $64,048$ annotated images. During the training process, we employ the classical instance segmentation technique, Yolact [19]. The annotation dataset is divided into two subsets: a training subset containing $57,644$ images and a validation subset containing $6,404$ images. Evaluation on the validation subset demonstrates that Yolact achieves an impressive accuracy of 93.08%(mAP) at a speed 36.3FPS on NVIDIA-A4000.

After extracting the target markers, precise positions of both the markers and the container need to be determined. In this paper, we designate the corner point of the marker (as indicated by white points in Fig. 3) and the corner points of the container (illustrated by the yellow points in Fig. 3) to denote their respective locations. As depicted in Fig. 1, four cameras are affixed to the spreader facing downwards. Consequently, the positions of the container's corner points

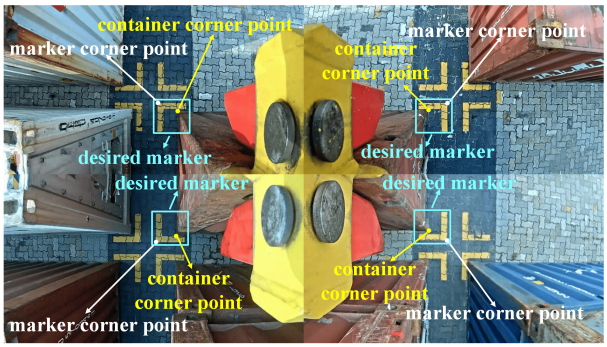
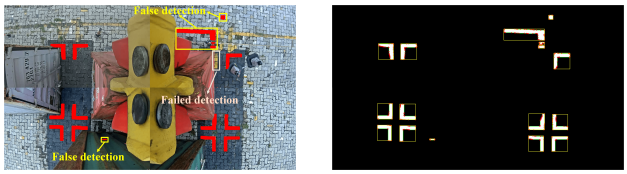


Fig. 3. The corner points of the markers and the container

remain constant in the images. Hence, we ascertain the positions of the container's corner points by manually selecting the corresponding corner pixels in the image. On the contrary, localizing the marker corner points is much more challenging. Consider, for instance, the scenario described in Fig. 4(a), where each camera captures a few target markers. Despite the utilization of deep learning techniques, achieving a flawless detection rate of 100% remains unattainable due to inherent uncertainties such as false or failed detection in foreground segmentation. Therefore, the primary challenge of this module lies in extracting the desired target markers from each image, while ensuring the absolute reliability of the extraction results in this safety-critical application.



(a) An example of false detection and failed detection (b) The corresponding binary image of the masks

Fig. 4. An example of failed foreground segmentation

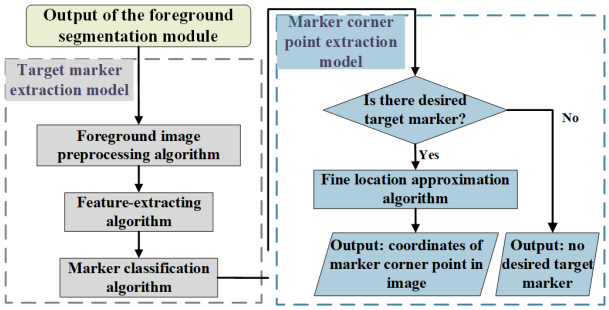


Fig. 5. The flow chart of the marker fine localization module

To address this challenge, we introduce a heuristic post-processing algorithm, which is composed of a target marker extraction model and a marker corner point extraction model (as outlined in Fig. 5). The details are provided as follows:

1) Target marker extraction model: In port operations, all unobstructed target markers share identical size and shape

characteristics, differing only in orientation, thus termed as standard target marker (as shown in Fig. 3 in this paper). These standard markers exhibit key features including an L shape, specific aspect ratio, defined area, and variable orientation, serving as the basis for designing feature extractors. Within this model, only the standard target markers are recognized as positive samples, whereas falsely detected or obstructed markers are categorized as negative samples. The foreground image preprocessing algorithm binarizes masks generated by the foreground segmentation module (as depicted in Fig. 4(b)). Subsequently, the canny edge detection algorithm [20] is applied to the binary image to delineate edges (depicted as green lines in Fig. 4(b)), enabling the determination of bounding boxes (illustrated as yellow rectangles in Fig. 4(b)). In practice, each mask edge constitutes a set of points (depicted as red points in Fig. 4(b)).

The feature extraction algorithm, detailed in Algorithm 1, consists of four steps. Firstly, four rectangular boxes are generated at the corners of each mask's bounding box (as depicted in Fig. 6). Secondly, a check is performed to ascertain if each rectangular box contains at least one edge point. For example, as illustrated in Fig. 6, if rectangular box 3 lacks any edge points while the remaining three contain at least one, it indicates the presence of an L-shaped target marker within this bounding box, oriented towards the lower right. Similarly, markers with different orientations can be identified. Thirdly, the aspect ratio of the mask's bounding box is computed by $r = \frac{w}{h}$, where w and h denote the width and height of bounding box, respectively. Finally, the area of the bounding box is determined by $A = w \times h$.

Algorithm 1 Feature extraction algorithm

Input: The bounding boxes and edges of n masks

Output: Features of all masks: $F = \{\boldsymbol{f}_i|_{i=1}^{n}\}$, $\boldsymbol{f}_i = \{L_i, r_i, A_i\}$, where L_i indicates the L shape and orientation feature, r_i is the aspect ratio, and A_i is the area

1: **for** $i = 1$; $i < n$; $i++$ **do**
2: Analyze shape and orientation: L_i (0: not L-shaped; 1: L shape with lower-right orientation; 2: L-shape with lower-left orientation; 3: L shape with upper-right orientation; 4: L shape with upper-left orientation)
3: Compute aspect ratio: $r_i = \frac{w_i}{h_i}$
4: Compute area: $A_i = w_i \times h_i$
5: **end for**
6: **return** F

In the marker classification algorithm, each mask is classified as either a positive sample or negative sample. To be classified as a positive sample, a mask must meet specific criteria regarding three features: L shape, aspect ratio, and area. In the proposed approach, we define the acceptable aspect ratio range as $[0.7,\ 1.3]$ and set the area threshold at 600 pixels.

2) Marker corner point extraction model: After extracting the target markers, the corner points need to be identified (as shown in Fig. 3). To achieve this objective, a fine location approximation algorithm (Algorithm 2) is developed. Firstly,

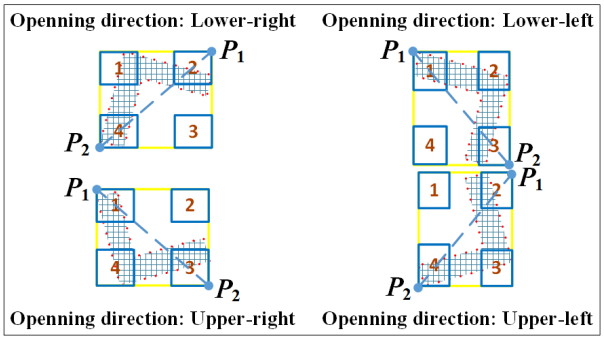


Fig. 6. The illustrative diagram of the marker corner point extraction module

as illustrated in Fig. 6, for positive samples oriented to the lower right or upper left, we connect the upper-right and the lower-left vertices of the bounding box in a straight line. Then we identify the edge point located above this line, farthest away from it, representing the corner point of the detected target markers. Similarly, corner points of target markers oriented to the lower-left or upper-right are determined using analogous procedures.

Algorithm 2 Fine location approximate algorithm

Input: Features F, bounding boxes with $[x_{tl}\ y_{tl}]^\top$, w, and h as the upper-left point, the width, and the height, respectively, edge points $\{\boldsymbol{p}_{eij}|_{j=1}^{m_i}\}$, where m_i is the number of edge points in the ith mask

Output: Marker corner points

1: **for** $i = 1$; $i < n$; $i++$ **do**
2: **if** L_i == 1 or L_i == 4 **then**
3: Compute diagonal $P_{1i}P_{2i}$, where upper-right point $P_{1i} = [x_{tli} + w\ y_{tli}]^\top$ and lower-left point $P_{2i} = [x_{tli}\ y_{tli} + h]^\top$
4: **else if** L_i == 2 or L_i == 3 **then**
5: Compute diagonal $P_{1i}P_{2i}$, where upper-left point $P_{1i} = [x_{tli}\ y_{tli}]^\top$ and lower-right point $P_{2i} = [x_{tli} + w\ y_{tli} + h]^\top$
6: **end if**
7: **for** $j = 1$; $j < m_i$; $j++$ **do**
8: **if** L_i == 1 or L_i == 2 **then**
9: Find the edge point above and with the largest distance to the diagonal $P_{1i}P_{2i}$: $\boldsymbol{p}_{ci}$
10: **else if** L_i == 3 or L_i == 4 **then**
11: Find the edge point below and with the largest distance to the diagonal $P_{1i}P_{2i}$: $\boldsymbol{p}_{ci}$
12: **end if**
13: **end for**
14: **end for**
15: **return** marker corner points: $\{\boldsymbol{p}_{ci}|_{i=1}^{n}\}$

B. Alignment Control

With the positions of both the container corner points and marker corner points determined, the relative pose of the container with respect to the target can be established.

Here, we denote $[\Delta x\ \Delta y]^\top$ and $\Delta\theta$ as the relative position and orientation of the container with respect to the target, respectively. As illustrated in Fig. 7, we stitch images from the four cameras and define a coordinate frame at the upper-left vertex of the resulting stitched image. The marker corner points of the desired target markers are defined as D_1, D_2, D_3 and D_4, while the container corner points are denoted as D_5, D_6, D_7 and D_8. Additionally, the center points of the container and the target are defined as D_9 and D_{10}, respectively. The positions of D_9 and D_{10} can be determined by

$$\begin{aligned} x_9 &= \frac{x_5+x_6+x_7+x_8}{4}, y_9 = \frac{y_5+y_6+y_7+y_8}{4} \\ x_{10} &= \frac{x_1+x_2+x_3+x_4}{4}, y_{10} = \frac{y_1+y_2+y_3+y_4}{4} \end{aligned} \tag{1}$$

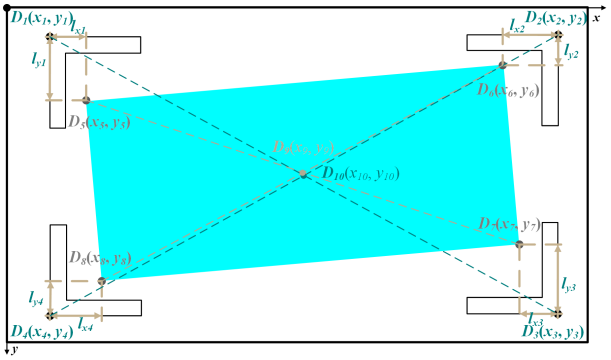


Fig. 7. The illustrative diagram of Δx, Δy, $\Delta\theta$

The relative position of the container with respect to the target can be computed as

$$\Delta x = x_{10} - x_9 = \frac{(l_{x2}-l_{x1})+(l_{x3}-l_{x4})}{4} \tag{2}$$

where $l_{x1} = x_5 - x_1$, $l_{x2} = x_2 - x_6$, $l_{x3} = x_3 - x_7$, and $l_{x4} = x_8 - x_4$.

$$\Delta y = y_{10} - y_9 = \frac{(l_{y4}-l_{y1})+(l_{y3}-l_{y2})}{4} \tag{3}$$

where $l_{y1} = y_5 - y_1$, $l_{y2} = y_6 - y_2$, $l_{y3} = y_3 - y_7$, and $l_{y4} = y_4 - y_8$. The relative orientation of the container with respect to the target can be computed as

$$\Delta\theta = \frac{\theta_1+\theta_2}{2} = \frac{\arcsin\frac{l_{y2}(l_{y1}-l_{y2})}{L_c l_{y1}} + \arcsin\frac{l_{y4}(l_{y3}-l_{y4})}{L_c l_{y3}}}{2} \tag{4}$$

where $\Delta\theta_1$ and $\Delta\theta_2$ represent the angles between two long sides of the container and the target, respectively, and L_c is the length of container's long side.

In practice, the deviation of the container and the target tend to be minimal. Thus, the disparities between l_{y1} and l_{y2}, as well as between l_{y3} and l_{y4}, are negligible. Under such circumstances, for computational simplicity, Eq. (4) can be approximated as

$$\Delta\theta \approx k\frac{(l_{y1}-l_{y2})+(l_{y4}-l_{y3})}{2} \tag{5}$$

where k is a constant parameter.

Having determined the deviation between the container and the target, the next step involves developing a control algorithm to minimize this deviation, ensuring precise alignment for the loading operation. As a specialized mechatronic system, the container crane is equipped with a micro-motion mechanism on the spreader to fine-tune its position and attitude, as illustrated in Fig. 8. Linear actuators 2 and 4 are tasked with moving Δx in synchronized motion. By coordinating the simultaneous movement of linear actuators 2 and 4 in the same direction, micro-translation along the gantry direction can be achieved. Similarly, simultaneous movement of linear actuators 1 and 3 in the same direction facilitates micro-translations along the trolley direction. Moreover, by controlling linear actuators 1 and 3 to move in opposite directions, micro-rotation of the spreader can be achieved. The schematic diagram detailing the control of linear actuators 2 and 4 is provided in Fig. 9(a), while the control mechanism for linear actuators 1 and 3 is elucidated in Fig. 9(b).

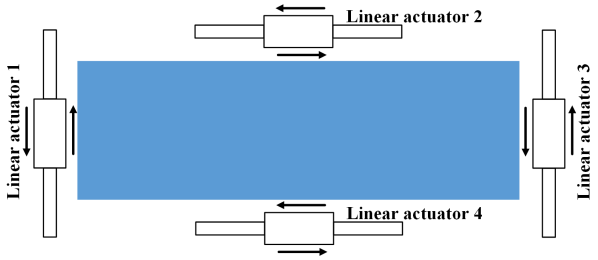
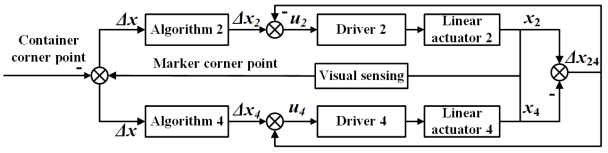
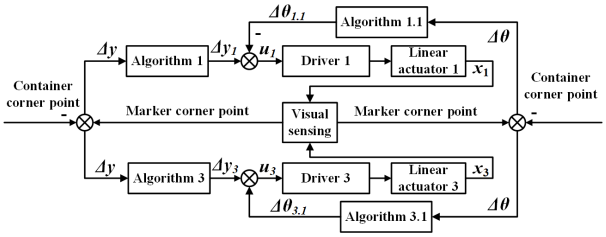


Fig. 8. The simplified schematic diagram of the micro-motion mechanism



(a) The control block diagram of linear actuators 2 and 4



(b) The control block diagram of linear actuators 1 and 3

Fig. 9. The block diagram of the proposed alignment control algorithm

As depicted in Fig. 9(a), two control objectives are outlined along the gantry direction. Firstly, linear actuators 2 and 4 are tasked with moving by Δx to minimize deviation towards zero. Secondly, the movements of actuators 2 and 4 should be synchronized. Notably, the unit of deviation Δx is expressed in pixels, while the displacements x_2 and x_4 of linear actuators 2 and 4 are in millimeters. Therefore,

algorithms 2 and 4 are implemented to translate Δx into Δx_2 and Δx_4, respectively, to correspond with x_2 and x_4. Similarly, the control objectives for actuators 1 and 3 are outlined in Fig. 9(b). Firstly, linear actuators 1 and 3 should move by Δy. Additionally, the movements of actuators 1 and 3 should induce the spreader to rotate by $\Delta\theta$. Here, the units of deviation Δy and $\Delta\theta$ are expressed in pixels, while the displacements x_1 and x_3 of linear actuators 1 and 3 are in millimeters. Algorithms 1 and 3 are utilized to translate Δy into Δy_1 and Δy_3, respectively, while algorithms 1.1 and 3.1 are employed to translate $\Delta\theta$ into $\Delta\theta_{1.1}$ and $\Delta\theta_{3.1}$, adapting to x_1 and x_3. This comprehensive approach ensures effective control, enabling the system to promptly respond to deviations and achieve synchronized motion across these three axes. It is worth noting that all control algorithms mentioned above are PID controllers.

IV. Experimental Results

To validate the effectiveness of the proposed approach in practical scenarios, a series of experiments were conducted on a real RTG crane at the Port of Ningbo, China. Firstly, the performance of target detection is discussed. The results are shown in Fig. 10. Masks enclosed by pink bounding boxes represent negative samples, whereas those enclosed by blue or red bounding boxes denote positive samples. In particular, the area of negative sample 1 is 546 pixels, falling below the 600 threshold. Negative samples 2, 3, and 4 fail to meet the L shape feature criterion, while negative sample 5 does not fulfill the aspect ratio criterion. Masks outlined in red bounding boxes exhibit the desired orientation. Extensive experiments were conducted to evaluate the accuracy of target marker extraction across various illumination and weather conditions, accounting for occlusions, damaged target markers, and image noise. The proposed target detection module underwent validation through $4,504$ image tests, achieving 100% accuracy rate based on human inspection of each image. This demonstrates the efficacy of the proposed heuristic postprocessing algorithm in overcoming the challenges posed by traditional CNN-based foreground segmentation algorithms. Through multi-feature verification, the proposed heuristic postprocessing algorithm effectively addresses uncertainties introduced by neural networks, rendering it capable of consistently providing reliable detection for the RTG control system. Selected experimental results are showcased in Fig. 11. Additionally, Fig. 11(b) illustrates that despite occlusions leading to failed detection in foreground segmentation, the proposed postprocessing algorithm successfully issued a warning signal ("No expected target"), ensuring the safe operation of the crane.

Next, the performance of alignment control is discussed. As depicted in Fig. 12, the relative deviation between the container and the target illustrates that the three components Δx, Δy, and $\Delta\theta$ all converge to the neighborhood of zero. This observation suggests that the proposed alignment control algorithm effectively aligned the container with the target markers. Additionally, Fig. 13 illustrates the control commands of four linear actuators, which likewise converge

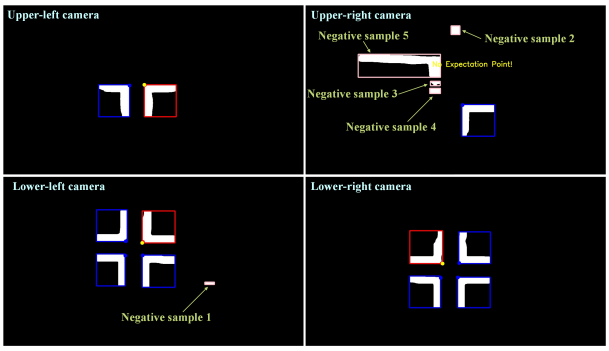
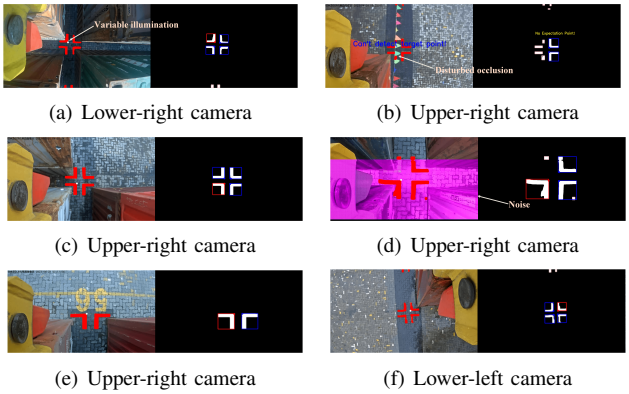


Fig. 10. Experimental results of target detection [The original images are shown in Fig. 4(a)]



(a) Lower-right camera (b) Upper-right camera

(c) Upper-right camera (d) Upper-right camera

(e) Upper-right camera (f) Lower-left camera

Fig. 11. Experimental results of target detection in special cases [(a) variable illumination; (b) disturbed occlusion; (c) damaged target marker; (d) noisy image; (e) cloudy; (f) rainy]

to the neighborhood of zero. The time complexity of the entire algorithm is approximately 25 FPS on an NVIDIA A4000 GPU and an Intel i7-13700K CPU. To reduce the impact of time variability caused by encoding and decoding during image transmission on the entire algorithm's processing speed, we proactively limited the algorithm's processing time to 10 FPS, which meets the requirements of the control system.

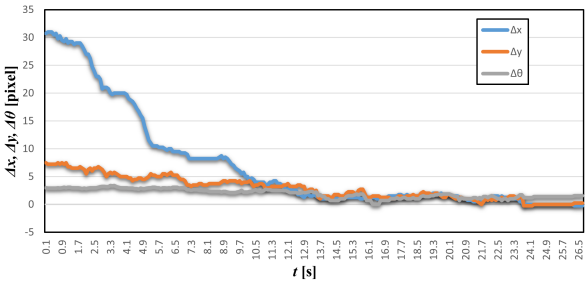


Fig. 12. The relative pose of the container with respect to the target [Δx, Δy, and $\Delta\theta$]

To demonstrate the superiority of the proposed approach over human operators, comparative experiments were conducted. A professional RTG operator operated the same RTG used to test the proposed approach. Manual and automated container loading operations were performed 10 times each.

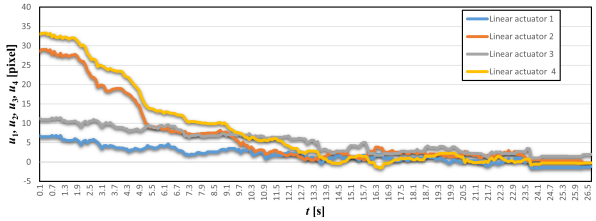


Fig. 13. The control commands of four linear actuators in the micro-motion mechanism

TABLE I

THE COMPARATIVE RESULTS

No.	Time $[s]$		Accuracy					
			Δx $[mm]$		Δy $[mm]$		$\Delta\theta$ $[mm]$	
	human	our	human	our	human	our	human	our
1	44.8	32.3	10	7	10	8	14	10
2	40.3	30.0	8	6	10	7	10	8
3	38.4	31.5	12	5	13	8	16	9
4	45.6	29.5	9	6	10	8	12	10
5	42.7	33.7	11	5	11	6	15	6
6	43.6	32.1	13	4	12	7	15	9
7	40.1	34.5	12	5	13	6	14	9
8	44.9	30.3	9	3	11	5	11	7
9	41.2	31.9	11	6	14	5	12	8
10	43.4	32.3	14	5	13	7	14	7
Average	42.5	**31.8**	10.9	**5.2**	11.7	**6.7**	13.3	**8.3**

The accuracy was evaluated by manually measuring the container position relative to the target at the final stage. The comparative results are presented in Tab. I, revealing that our proposed approach outperformed professional human operator in both accuracy and speed.

V. CONCLUSIONS

This paper has introduced an integrated approach for container loading of RTG cranes utilizing only local measurements from onboard sensors, with tight coupling between target detection and alignment control. The proposed approach relies on four monocular cameras mounted on the spreader, eliminating the need for additional sensors such as LiDAR or radar and showcasing its cost-effectiveness. By incorporating a heuristic image postprocessing algorithm for verification, the proposed approach ensures the correctness of detection results, thus ensuring the safe operation of these heavy-duty machines. Furthermore, it eliminates the requirement for calibrations of onboard sensors, streamlining the system and minimizing the sources of error. Successful testing on real RTG cranes validates the efficacy, robustness, and practicality of the proposed approach in industrial applications. This paper specifically addresses the loading containers on the yard ground. Future research will focus on container stacking to develop a comprehensive technical solution for container handling in the yard.

ACKNOWLEDGMENT

The majority of the work presented in this paper was conducted when the first author worked as a Postdoctoral Researcher at Shanghai Zhenhua Heavy Industries Co., Ltd. (ZPMC).

REFERENCES

[1] M. Hervás-Peralta, S. Poveda-Reyes, G. D. Molero, F. E. Santarremigia, and J.-P. Pastor-Ferrando, "Improving the performance of dry and maritime ports by increasing knowledge about the most relevant functionalities of the terminal operating system (tos)," *Sustainability*, vol. 11, no. 6, p. 1648, 2019.

[2] K. H. Kim and H. Lee, "Container terminal operation: current trends and future challenges," *Handbook of Ocean Container Transport Logistics: Making Global Supply chains Effective*, pp. 43–73, 2014.

[3] T. A. Sitompul, "Human–machine interface for remote crane operation: A review," *Multimodal Technologies and Interaction*, vol. 6, no. 6, p. 45, 2022.

[4] J. Benkert, R. Maack, and T. Meisen, "Chances and challenges: Transformation from a laser-based to a camera-based container crane automation system," *Journal of Marine Science and Engineering*, vol. 11, no. 9, p. 1718, 2023.

[5] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[6] E. E. Hemayed, "A survey of camera self-calibration," in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2003.* IEEE, 2003, pp. 351–357.

[7] L.-I. L. GmbH, "Truck positioning crane," 2022. [Online]. Available: https://lase-solutions.com/wp-content/uploads/2022/03/LaseTPS_-_Truck_Positioning_System.pdf

[8] SIEMENS, "Simocrane truck positioning system (tps)-highly precise laser measurement system for accurate truck positioning," 2023. [Online]. Available: https://assets.new.siemens.com/siemens/assets/api/uuid:6286d728-87e9-4c13-9ada-7ae32e282e07/vrtl-b10009-00-7600-144dpi-simocrane-truck-positioning-system.pdf

[9] Siemens, "Simocrane final landing system (fls)," 2023. [Online]. Available: https://assets.new.siemens.com/siemens/assets/api/uuid:31c62b34-5ec2-436a-8b51-dad9fa14ab0d/vrtl-b10019-00-7600-144.pdf

[10] P. Blaiklock, "Automated stacking cranes," 2023. [Online]. Available: https://wpassets.porttechnology.org/wp-content/uploads/2019/05/25183601/052-053_3.pdf

[11] H. Kawai, Y. Choi, Y. Kim, and Y. Kubota, "Position measurement of container crane spreader using an image sensor system for anti-sway controllers," in *2008 International Conference on Control, Automation and Systems*. IEEE, 2008, pp. 683–686.

[12] H. Kawai, Y.-B. Kim, and Y. Choi, "Measurement of a container crane spreader under bad weather conditions by image restoration," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 1, pp. 35–42, 2011.

[13] J.-J. Lee, G.-G. Nam, B.-K. Lee, and J.-M. Lee, "Measurement of 3d spreader position information using the ccd cameras and a laser distance measuring unit," *Journal of Navigation and Port Research*, vol. 28, no. 4, pp. 323–331, 2004.

[14] J. J. Lee, G. G. Nam, B. K. Lee, and J. M. Lee, "Measurement of 3d spreader position for automatic landing system," in *30th Annual Conference of IEEE Industrial Electronics Society, 2004. IECON 2004*, vol. 3. IEEE, 2004, pp. 3008–3013.

[15] H.-J. Yoon, Y.-C. Hwang, and E.-Y. Cha, "Real-time container position estimation method using stereo vision for container auto-landing system," in *ICCAS 2010*. IEEE, 2010, pp. 872–876.

[16] Y. Liu, T. Li, L. Jiang, and X. Liang, "Container-code recognition system based on computer vision and deep neural networks," in *AIP Conference Proceedings*, vol. 1955, no. 1. AIP Publishing, 2018.

[17] N. Hütten, R. Meyes, and T. Meisen, "Vision transformer in industrial visual inspection," *Applied Sciences*, vol. 12, no. 23, p. 11981, 2022.

[18] C. Tang, P. Chen, and Y. Li, "Automatic damage-detecting system for port container gate based on ai," in *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition*, 2020, pp. 146–151.

[19] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "Yolact: Real-time instance segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9157–9166.

[20] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 679–698, 1986.