

Received August 25, 2020, accepted September 4, 2020, date of publication September 9, 2020, date of current version September 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3022906

# A Workpiece Localization Method for Robotic De-Palletizing Based on Region Growing and PPHT

JINQUAN LI<sup>1</sup>, (Member, IEEE), JIA KANG<sup>1</sup>, ZHE CHEN<sup>2</sup>, FANG CUI<sup>1</sup>, AND ZIXUAN FAN<sup>1</sup>

<sup>1</sup>School of Modern Posts, Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>2</sup>Institute of Engineering Medicine, Beijing Institute of Technology, Beijing 100081, China

Corresponding authors: Jinquan Li (lijinquan@bupt.edu.cn) and Zhe Chen (zhechen95@qq.com)

**ABSTRACT** Aiming at the problem that the robot de-palletizing task is difficult to accomplish under unstable ambient light, a two-step method is proposed to realize the localization of workpieces, which in this work are woven bags. To begin with, Region Growing method is used to extract the whole target region in the original image, and the relationship model between image intensity and the optimal Region Growing threshold is established. Then, Progressive Probabilistic Hough Transform (PPHT) is used to locate each woven bag. To improve the system performance, the optimal parameters of the PPHT function in different illumination intervals are determined. Finally, experiments are conducted to verify the effectiveness of the proposed method. Experiment results demonstrate this method is robust and feasible.

**INDEX TERMS** Image segmentation, workpiece localization, region growing, PPHT.

## I. INTRODUCTION

On automated production lines, palletizing and de-palletizing job is an important link connecting production and transportation [1]–[3]. In order to improve the production efficiency of this link, de-palletizing robots came into being, which can firmly grasp and deliver goods with a special customized multi-functional grasper. With the development of industrial robots and industrial control technology, de-palletizing robots are gaining prevalence in diverse industries [4]–[6]. However, traditional industrial robots can hardly handle the grasping tasks at complex scenes when the objects are not fixed accurately, because offline programming or “teaching and playback mode” are heavily relied on [7]. To improve the generalization ability of industrial robots, aiding robots with machine vision is becoming increasingly prevalent [8]–[11].

At present, machine vision technology is widely used to detect the appearance and quality of products and packages [12]. Recently, several studies on machine vision-based product positioning and packaging in robotic de-palletizing operations are reported. Rahardja’s research used stereo cameras [13] to calculate the location of random stacked components. However, the system requires unique landmark features to identify the target object and estimate the pose of

the object. Guo Jin [14] proposed a fast part hole distance detection technique based on a binocular vision sensor, which can accurately locate the three-dimensional pose of the part. However, the binocular vision system needs to decode and match the images taken by both cameras, so the matching error will affect the accuracy of recognition and positioning. Xinjian Fan [15] developed a prototype automatic palletizing system that combines industrial robots with binocular stereo vision. The shape-based matching (SBM) method was used instead of traditional stereo matching in this method to improve the performance of three-dimensional pose estimation of a workpiece with high robustness. However, the system needs to pre-register the pattern of the tested object, and then train the model according to the key features of the object. The final recognition accuracy is greatly affected by the model. Recently, there have also been studies focusing on the introduction of structured light sensors [16]. In these methods, the attitude estimation is achieved by comparing the three-dimensional model of the workpiece with the distance image [17] or the three-dimensional point cloud [18] captured by the structured light sensor. However, the above methods are not suitable for the bad conditions of surface reflection or light interference. Zhang Biao [19] proposed a new method of non-calibrated vision based on space operation and three-dimensional laser-assisted detection for locating targets. The system uses laser spot recognition to detect

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

objects, which can avoid the influence of the illumination environment. However, the system is complex and the number of projected laser points needs to vary according to the size of the workpiece. And we find many interference factors exist under practical complicated working environment, which complicates the identification and localization of single woven bag: 1) the ambient light varies with the weather, time, human-induced disturbance among others, which may cause image intensity varying intensely; 2) background interference, which includes but is not limited to shelves, conveyor belts and robot body, blends target with inseparable noises; 3) with diverse palletizing types and deformable objects, the flexible edges of adjacent woven bags overlap with each other. With all these obstacles under consideration, we selectively ascertain the superior existing method and finely tune the optimal parameters with our own adaptive model. Our algorithm is designed explicitly to tackle the demanding practical environment, and more importantly to achieve superior identification results.

In this paper, a two-step recognition method is proposed. The Region Growing method is selected to extract the whole stack, by considering the adhesive characteristics of the edge of the woven bag. Then a linear model between the optimal threshold of regional growth and the difference between seed point and average brightness is established, through which the edges of each woven bag are obtained by combining the adaptive threshold segmentation and PPHT. Contrary to the traditional edge detection method, the proposed one can obtain more detailed contour information. The corresponding relationship between the three parameters combinations of PPHT and the illumination interval is also analyzed to enhance the robustness of the system to illumination variation. Finally, the pose information of a single workpiece is calculated to guide the robot to de-palletize. With the proposed algorithm, we have addressed the actual industrial problem that may strike us all, with low cost and high applicability, and cast light for generalizing our methodology to other conventional algorithms.

The remaining of this paper is organized as follows: In Section II, the composition of the visual acquisition system is described. In Section III, several image segmentation algorithms are compared for the whole stack extracting task, from which Region Growing is chosen for its superior results. After that, under image average intensity, a model between image average brightness and the optimal growing threshold is established. In Section IV, the optimal parameters of the PPHT function for different image intensity intervals are determined to locate a single woven bag. In Section V, experiments are conducted to verify the feasibility of the proposed method, and detailed discussion on the results is presented. In Section V, conclusions are drawn.

## II. VISION SYSTEM DESIGN

In this work, an image acquisition system for real industrial application is established. It mainly includes a Basler Gigabit Ethernet industrial camera acA3800-10gm, a laser

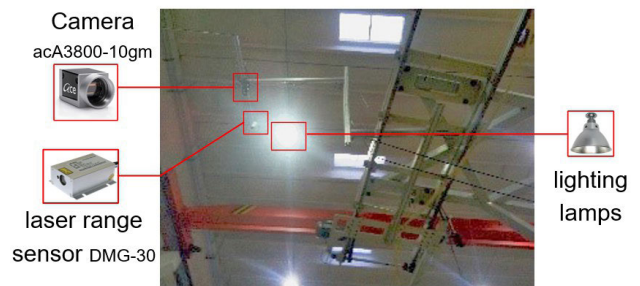


FIGURE 1. Image acquisition system.

range sensor DMG-30, and illumination devices, in which the camera and the range sensor are fixed, and the range sensor is perpendicular to the ground to obtain the depth information of the workpiece. The camera has a resolution of  $3856 \times 2764$  pixels. The visual acquisition system is shown in Fig. 1.

Considering that the woven bags are prone to deform, this paper introduces a general method, taking the stack of single or multi-layer woven bags as the research object. The main palletizing types used in this experiment are criss-crossing type and forward-reverse type, rather than overlap type and rotating crisscross type. Since the overlap type features instability between layers, and rotating crisscross type forms holes in the middle of the stack, which reduces the utilization rate of the pallet.

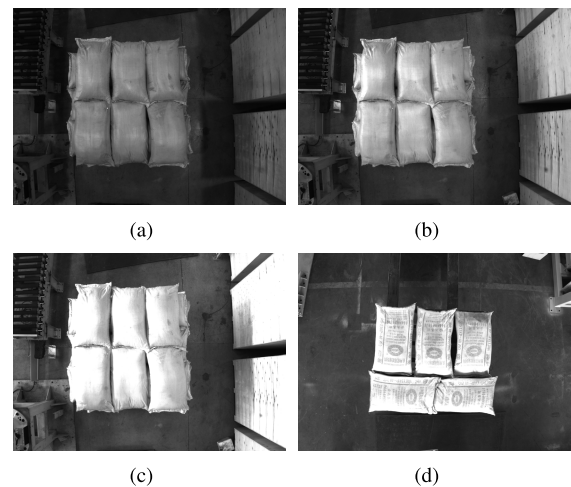


FIGURE 2. Sample image under different illumination. (a) Criss-crossing type,  $m=59.92$ . (b) Criss-crossing type,  $m=90.31$ . (c) Criss-crossing type,  $m=128.78$ . (d) Forward-reverse type,  $m=90.89$ .

As shown in Fig. 2, the sample images with different de-palletizing types are acquired by the above-mentioned visual acquisition system under different illumination circumstances, where  $m$  is the average intensity of the sample image. When the ambient illumination changes, the average intensity of the sample image collected varies accordingly, and its variation range is approximately 30-130 in this paper.

In this paper, we aim to identify and locate a single woven bag in the de-palletizing environment with unstable natural light. The edges of adjacent woven bags are generally

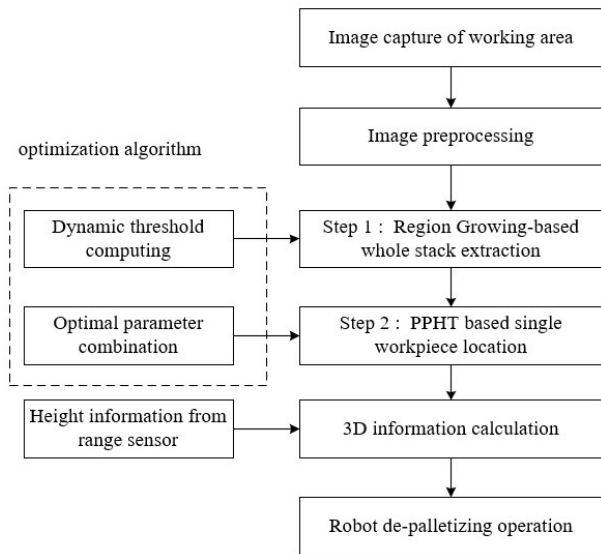


FIGURE 3. The system workflow.

obscure in collected images because of their highly deformative property. Therefore, it is hardly probable to recognize and locate a single woven bag directly. Furthermore, the original images need to be preprocessed due to the existence of interference in the actual de-palletizing environment such as shelves, conveyor belts and machines. To this end, a two-step recognition method is proposed in this paper. In the first step, the Region Growing method with the aid of an adaptive model is used to extract the whole stack. In the second step, the PPHT algorithm based on the optimal parameters model is used to extract a single woven bag. After that, the position and orientation of each woven bag are obtained. Finally, combined with the height information from the range sensor, the 3d coordinates of each workpiece are calculated to guide the robot to de-palletize. The architecture of the proposed system is shown in Fig. 2.

### III. WHOLE STACK EXTRACTION

In this section, the adaptive threshold, local threshold and Region Growing are compared for the whole stack extraction task, and the Region Growing method is selected for its superior performance. At the same time, considering the results of image segmentation will be influenced by natural light variation in the real environment, a relationship model between the average image intensity and the optimal threshold for Region Growing is established based on a large number of experiments. Based on this model, the optimal threshold can be ascertained to improve the accuracy of segmentation and relieve the impact of natural light instability on image segmentation.

#### A. IMAGE PREPROCESSING

The adaptive threshold, local threshold and Region Growing methods are performed for the whole stack extraction task. The results of Fig. 2b are shown in Fig. 4.

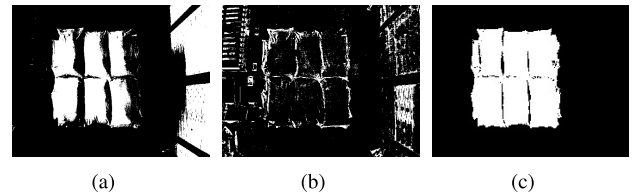


FIGURE 4. Results of (a) Adaptive threshold, (b) Local threshold, and (c) Region Growing.

As shown in Fig. 4a, the adaptive threshold method is not ideal for image segmentation with low contrast, and the right side of the woven bags is integrated with the background, leading subsequent processing impossible. In Fig. 4b, useful edges are submerged by noise, and the edge of the woven bag is not obvious. Therefore, the segmentation of woven bags cannot be achieved through the local threshold method because the image intensities of foreground and background don't necessarily differ a lot. In Fig. 4c, the ideal segmentation effect can be obtained through the Region Growing method with an appropriate threshold. Therefore, the Region Growing method is chosen to extract the whole stack for its superior results.

Region Growing method [20], [21] starts with initializing seeds and links the neighboring pixels according to growing formula until all pixels being labeled. The growing formula based on the gray difference between neighboring pixels and seeds can be expressed as:

$$|f(x, y) - S|_{(x,y) \in P} < K \quad (1)$$

In (1),  $S$  is the gray value of seed point,  $f(x, y)$  is the gray value of the neighboring pixels of seed point. Considering that different stacks are always located in an approximately fixed area in the images, the fixed target seed point can be chosen for Region Growing.

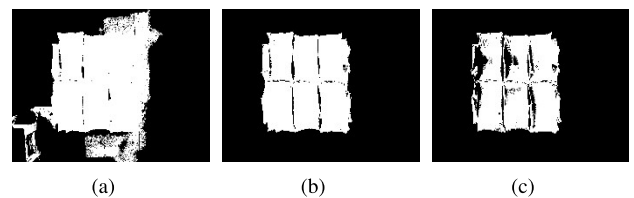


FIGURE 5. Segmentation results using fixed threshold-based Region Growing. (a) The segmentation result of Fig.2a. (b) The segmentation result of Fig.2b. (c) The segmentation result of Fig.2c.

#### B. DYNAMIC THRESHOLD

Although the Region Growing method performs excellently in extracting whole stack, the image growing method with the same threshold value cannot achieve desirable results. The segmentation results of Fig. 2a,b,c by using the Region Growing method with a fixed threshold ( $K=70$ ) are shown in Fig. 5a,b,c. In Fig. 5a, a part of the background is mistakenly segmented into the target pile, indicating that the threshold is too large for 2a with darker light intensity. In Fig. 5b,

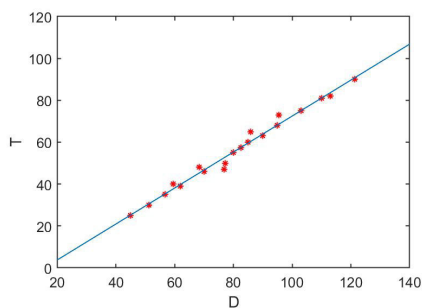
the target piles are completely identified and segmented, indicating that the threshold is appropriate for 2b with moderate light intensity. In Fig. 5c, some areas of the target pile are not completely extracted, indicating that the threshold is too small for 2c with strong light intensity. Therefore, the Region Growing based on a fixed threshold cannot extract desired stacks effectively under different illuminations.

Considering ambient illumination is mainly reflected in the average brightness of the image, we select the images under different illuminations for segmentation experiments, to find the relationship between the average brightness of each image and the optimal threshold for the Region Growing method.

**TABLE 1.** Region Growing parameter table of images under different illuminations.

Number	1	2	3	4	5
M	36.10	60.71	88.04	107.09	120.70
S	98.00	138.00	174.00	220.00	242.00
D	61.90	77.29	85.96	112.91	121.30
T	39.00	50.00	65.00	82.00	90.00

Table 1 gives some results under different illuminations, in which  $M$  is the average brightness,  $S$  is the gray value of seed point,  $D$  is the absolute value of the difference between  $S$  and  $M$ , and  $T$  is the optimal growing threshold. The corresponding ( $D$ ,  $T$ ) values of 21 images under different illuminations are calculated and drawn with the red discrete points in Fig. 6. A least-square algorithm fitting line demonstrates the linear relationship between  $D$  and  $T$ , which is shown by a blue line in Fig. 6.



**FIGURE 6.** ( $D$ ,  $T$ ) least squares fitting results.

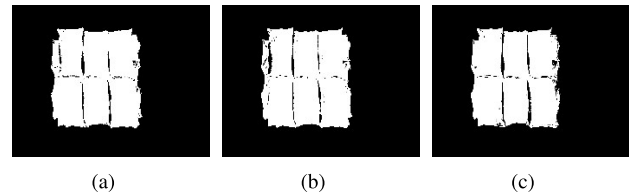
The equation for the fitted line is:

$$T = 0.8593 * D - 13.4723 \quad (2)$$

Therefore, the relationship between  $T$ ,  $M$  and  $S$  is:

$$T = 0.8593 * (|M - S|) - 13.4723 \quad 35 \leq M \leq 150 \quad (3)$$

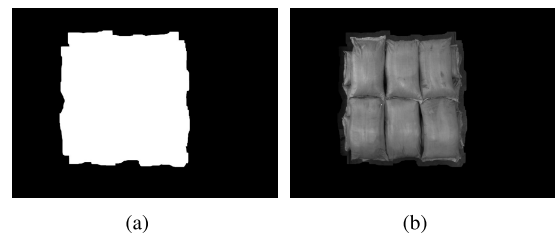
When  $M < 35$ , it indicates that the environment is too gloomy, so the light needs to be turned on; when  $M > 150$ , the ambient light is too intense, which rarely occurs in the general workshop. Equation (2) is the relationship model between image average intensity and optimal growing threshold. According to the model above, the optimal threshold for



**FIGURE 7.** Segmentation results using dynamic growing threshold. (a) The segmentation result of Fig.2a. (b) The segmentation result of Fig.2b. (c) The segmentation result of Fig.2c.

Region Growing of Fig. 2a,b,c can be determined as 50, 70, 94, respectively. And the results are shown in 7a,b,c. It can be seen that the target stack is completely segmented. Compared with Fig. 5, the effects of illumination are almost eliminated by using dynamic thresholds in Fig. 7. And the whole stack can be completely detected, while the single woven bag is hardly identified due to obscure edges of braided bags.

Considering that there may be some boundary points between adjacent woven bags and points between the stack and the background in Fig. 7a, a morphological expansion operation is performed on Fig. 7a to expand the growing area, leading to a more complete stacking area with reduced errors, as shown in Fig. 8a. And then an AND operation between the original image Fig. 2a and Fig. 8a is performed to get the stacking area on the original image, with the result shown in Fig. 8b. Apparently, through above image processing steps, the whole stack can be extracted accurately from background and influence of noise is greatly reduced, which lays the foundation for the accurate identification of subsequent single woven bag.



**FIGURE 8.** (a) Mask of extracted stacking area. (b) The stacking area presented on the original image.

#### IV. SINGLE WORKPIECE LOCATION

The purpose of this paper is to identify each woven bag in an actual de-palletizing environment under unstable natural light. In Section III, the whole stack has been extracted, removing most of the interference outside the target pile area. In this section, we will identify and locate the central position and orientation of each woven bag. Contour extraction result of a single woven bag is not satisfying and numerous informative boundaries are lost when applying traditional edge extraction algorithm, while adaptive threshold segmentation is capable of extracting more contour information and reduces the influence of intensity variation. Therefore, the adaptive threshold segmentation is used for edge points detection,

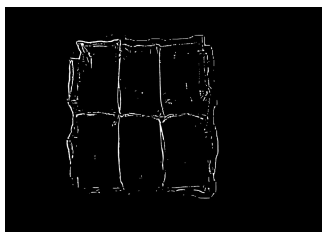


FIGURE 9. Edge extraction image.

as shown in Fig. 9, and then the optimal parameter adjustment is performed on the PPHT function. Finally, the pose information of each woven bag is obtained.

In the PPHT function, the main parameters encompassing  $Th$ (int threshold),  $Min$ (double min Line Length), and  $Max$ (double max Line Gap) primarily affect the line extraction effect, in which  $Th$  is the threshold of the number of points determining whether a line exists or not;  $Min$  is the threshold of the minimum length of a line; and  $Max$  is the threshold of the maximum gap between points on the same line to link them. From the analysis of the definition, the details detected will increase with these three parameters decrease, while excessive restoration of unnecessary details may lead to recognition failure. Therefore, this paper will analyze the appropriate PPHT parameters according to different light intensity values, to better segment and identify the contour of a single bag.

After roughly summarizing the test results of several pictures, it is found that the suitable parameters of PPHT did not change much for images when illumination intensity changed slightly, so the appropriate interval of image intensity is set as 10. Since the effective light intensity interval is (55, 125), we take those pictures with the brightness values close to 60, 70, 80, 90, 100, 110 and 120 as the representative sample images in each illumination interval. However, when  $Th$ ,  $Min$  and  $Max$  are greater than 400, the image cannot be effectively recognized, and slight variation in these parameters can hardly produce significant changes, so the three parameters are adjusted at intervals of 50, with values ranging from 50 to 350. As shown in Fig. 11(a-n), a total of 7 images are counted, and a total of 343 sets of data are counted for each image. The statistically preferable parameter results of each image are composed of two statistical charts: fix  $Th$  and  $Min$  respectively, and then count and analyze the remaining two parameters to find the optimal parameter combination.

The optimal parameter combination is determined by following steps. Take Fig. 11e and f as an example, which is a complete analysis for the average intensity being around 80. First, the parameter  $Max$  is set as 50, then all combinations of the other parameters are tested, and those efficient ones are recorded a dark blue point in the chart Fig. 11e. Then, the parameter  $Max$  is set as 100, ..., 350, respectively, and corresponding combinations of the other parameters are recorded in the same chart with points of different colors. After that, those combinations with the maximal times of

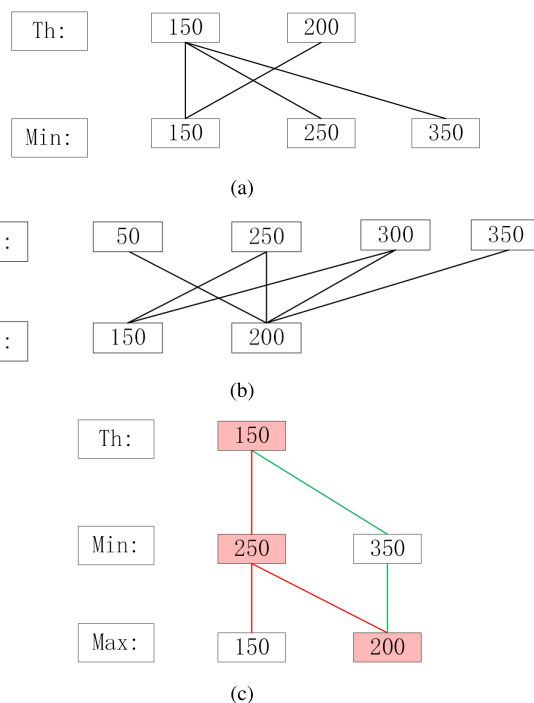
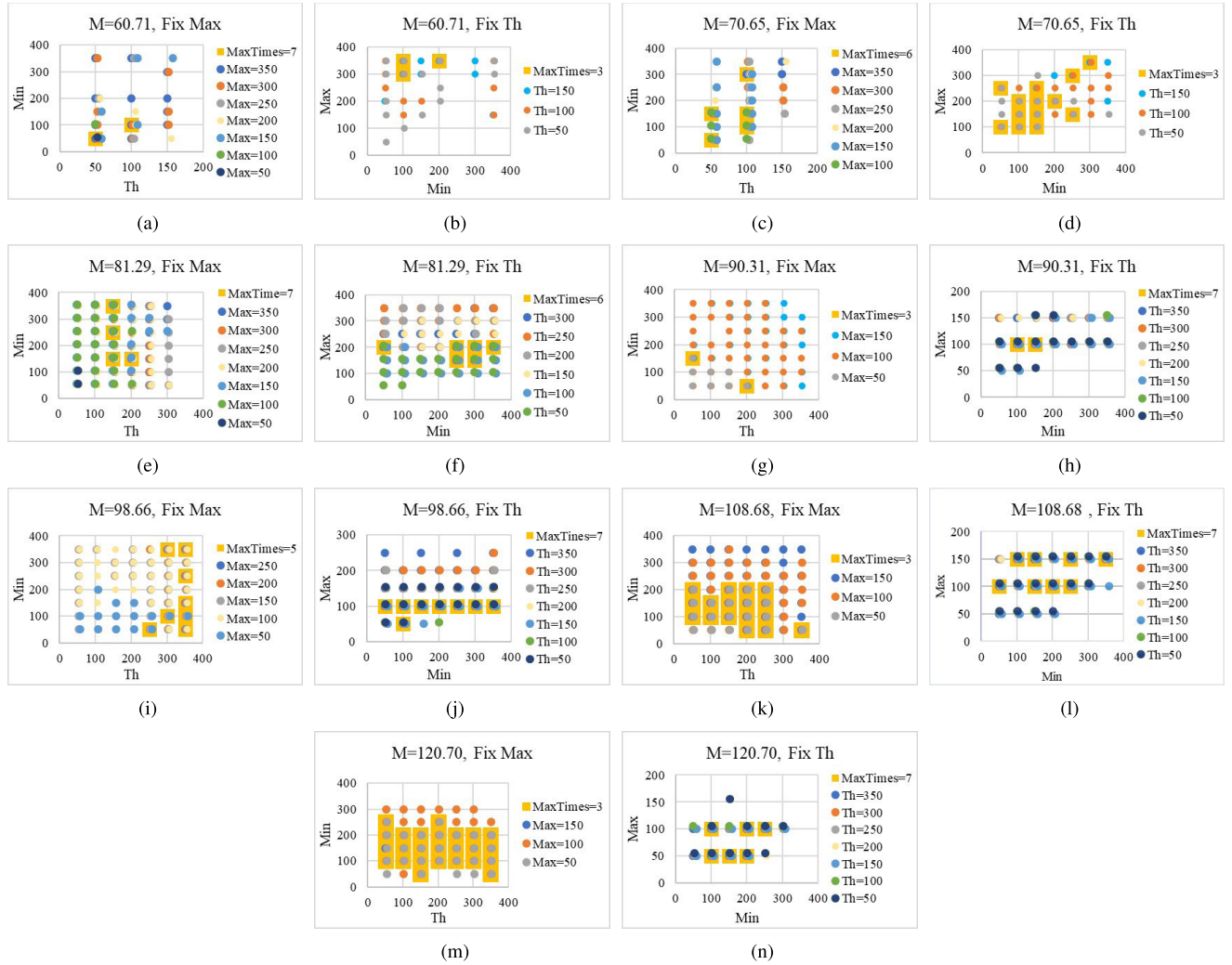


FIGURE 10. The most universal combinations when (a) fixing  $Max$ , and (b) fixing  $Th$  in one particular example. (c) Intersection of (a) and (b).

successful recognition are located and labeled with yellow square. In Fig. 11e, the maximal times of successful recognition is 7, which means no matter what  $Max$  is, these combinations of the other two parameters can suit our needs. In addition, in order to find the most universal combination of all the three parameters, the second parameter  $Th$  is set from 50 to 300, with the interval being 50. Then, as in the former step, those combinations of  $Max$  and  $Min$  with the maximal times of successful recognition are located and labeled with a yellow square in Fig. 11f. The successful combinations at chart Fig. 11e are shown in Fig. 10a, and successful combinations at Fig. 11f can be seen in Fig. 10b. By combining the former two figures, we can obtain Fig. 10c. Combination with  $(Th, Min, Max)$  being (150, 250, 200) is the most universal one because parameters with red block have more branches, which proves it to be the most robust one for generalizing other images.

After a series of experiments, analyses and further tests on other images, the optimal parameters corresponding to each illumination interval are obtained, as shown in Table 2.

In Fig. 12a, the contour of each woven bag in Fig. 9 is detected by using the optimal parameters given in Table 2. Due to the existence of pores between the extracted lines, we connect them by proper morphological closure operation to obtain better contour. The morphological dilation operation is performed on Fig. 12a to obtain Fig. 12b, and then the morphological erosion operation is performed on Fig. 12b to obtain Fig. 12c. Finally, the contour of each woven bag in Fig. 12d is obtained by extracting the inner contour from Fig. 12c.



**FIGURE 11.** The statistical preferable parameter results of images with different illumination. The parameter MaxTimes is the maximal times of successful recognition. (a)  $M=60.71$ , fixed Th. (b)  $M=60.71$ , fixed Max. (c)  $M=70.65$ , fixed Th. (d)  $M=70.65$ , fixed Max. (e)  $M=81.29$ , fixed Th. (f)  $M=81.29$ , fixed Max. (g)  $M=90.31$ , fixed Th. (h)  $M=90.31$ , fixed Max. (i)  $M=98.66$ , fixed Th. (j)  $M=98.66$ , fixed Max. (k)  $M=108.68$ , fixed Th. (l)  $M=108.68$ , fixed Max. (m)  $M=120.70$ , fixed Th. (n)  $M=120.70$ , fixed Max.

**TABLE 2.** Corresponding PPHT parameter values for illumination intervals.

interval	parameter		
	Th	Min	Max
55-65	100	100	350
65-75	100	150	200
75-85	150	250	200
85-95	50	150	100
95-105	350	100	100
105-115	250	150	100
115-125	350	200	100

After the contour of each woven bag is extracted, the circumscribed rectangle with a minimum area can also be obtained, whose center is exactly the central position of each workpiece [22]. Fig. 13b shows the extraction result of the single woven bag placed vertically on the first one of the

second row, in which the centroid of its circumscribed rectangle represented by a white point is exactly the central position of workpiece.

When the robot performs the de-palletizing job, in addition to the center position of each workpiece, the posture information of each workpiece is also required. Fig. 14 and Fig. 15 show the placement postures of the separated workpieces respectively, in which  $W$  is the first rectangular side parallel to the horizontal axis when the horizontal axis rotates counterclockwise, and the other side is denoted as  $H$ . The angle between line  $W$  and the horizontal axis is denoted as  $\vartheta$ , which is obtained through the circumscribed rectangle extracted. The posture of the woven bag is denoted as  $\beta$ , which is calculated by using (4). When the robot is grabbing a woven bag, it adjusts the grabbing direction according to the posture of the woven bag. The attributes of the workpiece in Fig. 13 are shown in Table 3.

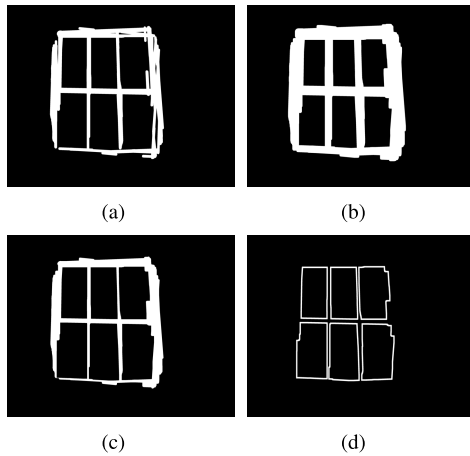


FIGURE 12. (a) The result of PPHT. (b)The result of morphological dilation operation. (c)The result of morphological erosion operation. (d) The contour of a single braided bag.

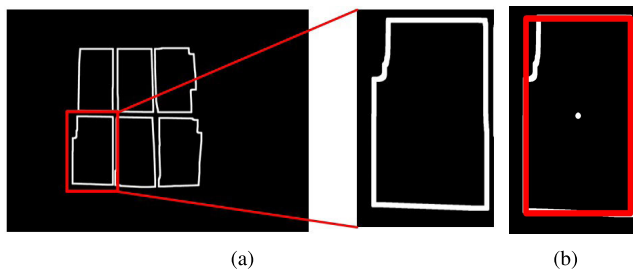


FIGURE 13. (a)The left woven bag of the second row and its enlarged view. (b)The fitted rectangle of this woven bag.

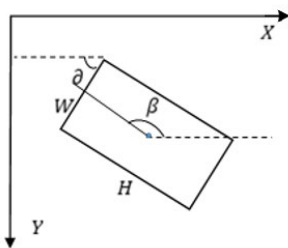


FIGURE 14. Workpiece placed with attitude 1.

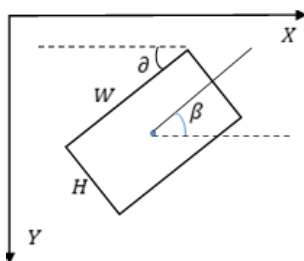


FIGURE 15. Workpiece placed with attitude 2.

$$\beta = \begin{cases} \theta & W \geq H \\ \theta + 90^\circ & W < H \end{cases} \quad (4)$$

After the central position and orientation information of each woven bag is obtained, combined with the depth

TABLE 3. Workpiece pose properties.

x/pixel	y/pixel	$\beta/^\circ$
1113	1749	90.00

information from the range sensor, the 3D coordinate information of each workpiece at the first layer can be obtained, and uploaded to the robot control system. Then the gripper is controlled to grab the corresponding workpieces at the first layer. Iteratively, the above procedures are repeated for the next layer, which doesn't give rise to any challenges not covered by the abovementioned algorithm. Layer-by-layer, the whole stack is de-palletized.

### V. EXPERIMENTS AND DISCUSSION

To prove the effectiveness of the vision system proposed in this paper, we built an actual robot de-palletizing job environment in Qingdao Baojia Automation Equipment Co., Ltd., as shown in Fig. 16. Three parts of experiments under the environment of unstable natural light are carried out to verify the applicability of the system to different de-palletizing types, the accuracy of the system to identify and locate the woven bag, the robustness of the system to the change of illumination, respectively.

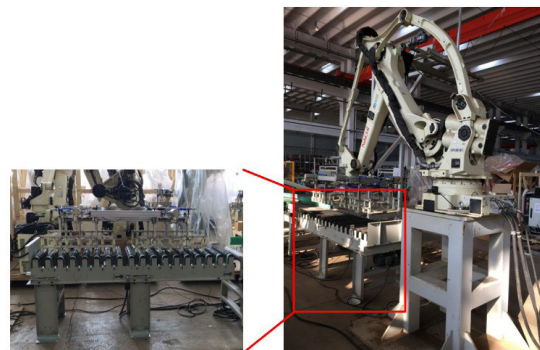
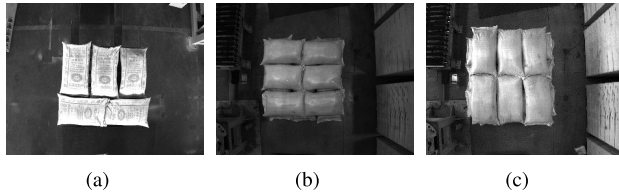


FIGURE 16. The field test photos.

#### A. THE APPLICABILITY OF THE SYSTEM TO DIFFERENT DE-PALLETIZING TYPES

To test the applicability of the system, we need to experiment with different de-palletizing types. we consider that the identification of woven bags is mainly for woven bag postures information. The major differences between different stack layers are the center position, orientation and size of each woven bag. In Section II, we introduce and analyze the advantages and disadvantages of several pallet-type commonly used in the industry. The criss-crossing type and forward-reverse type have well stability and wide application, and also cover the change of the positions and orientations in the type of horizontal, vertical or forward-reverse. Meanwhile, the stack of five-flower and six-flower per layer are

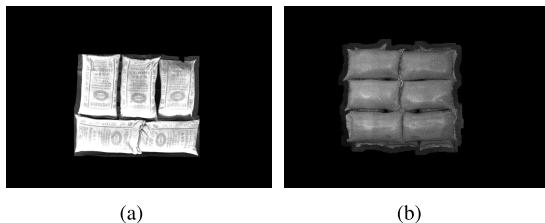


**FIGURE 17.** The images of the stack of different types: (a) five-flower stack, (b) six-flower horizontal stack, and (c) six-flower vertical stack.

common quantity in the industry, and more/less woven bags per layer can hardly bring about any new and neglected issues. Therefore, the experiments on the typical five-flower stack, six-flower horizontal stack and six-flower vertical stack, which cover the change in the positions and orientations of woven bags, and the number of woven bags per layer, can effectively demonstrate the feasibility of the proposed algorithm.

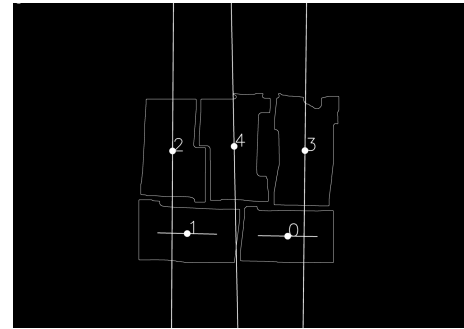
We collect 100 images in a natural light unstable environment at different times on sunny and cloudy days, which include five-flower stacks and six-flower horizontal or vertical stacks, as shown in Fig. 17a,b,c. As aforementioned, the localization of six-flower vertical stack has already been realized. To verify the applicability of the system, the experiments on the other two types are carried out.

In Fig. 18, the whole stacking area is completely extracted from Fig. 17a,b, by using the optimal segmentation threshold from the model constructed in Section III. Thus, with our method, the whole stacking area can be accurately extracted from the background for different illumination and stacking types.

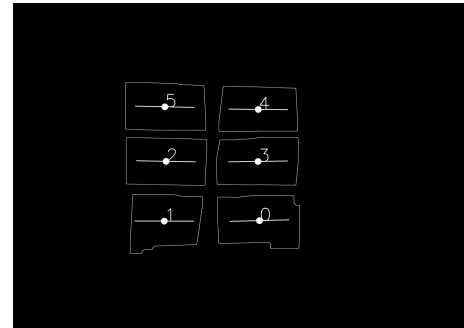


**FIGURE 18.** The complete stacking area of (a) five-flower stack, and (b) six-flower horizontal stack.

Then, the identification and localization results of each woven bag are obtained by using PPHT with the optimal parameters given in Table 2. In Fig. 19a,b, the center points of woven bags with different postures are marked with white points, and the number is marked for each woven bag. The center position coordinates and orientation information of each bag are shown in Tables 4 and 5. As shown in the results above, the proposed algorithm can extract with high accuracy the posture of each woven bag with a wide range of position, orientation and sizes (caused by layer height variation), proving the feasibility and applicability of the algorithm to different de-palletizing types under the environment of unstable natural light.



(a)



(b)

**FIGURE 19.** The result of identification and location of single woven bag in (a) five-flower stack, and (b) six-flower horizontal stack.

**TABLE 4.** The center position coordinates and orientation information for each woven bag in the five-flowered stacking.

Number	x/pixel	y/pixel	$\beta/^\circ$
0	2323	1971	179.43
1	1473	1949	178.85
2	1349	1251	89.38
3	2469	1248	89.38
4	1870	1212	91.19

**TABLE 5.** The center position coordinates and orientation information for each woven bag in the six-flowered stacking.

Number	x/pixel	y/pixel	$\beta/^\circ$
0	2084	1834	1.72
1	1279	1838	0.00
2	1294	1335	178.85
3	2072	1336	0.57
4	2074	897	178.43
5	1284	875	178.85

## B. THE ACCURACY OF THE SYSTEM TO IDENTIFY AND LOCATE THE WOVEN BAG

In order to evaluate the accuracy of identification of woven bag with unstable natural light, we analyze the error between the center point coordinates detected by the system and manually extracted. As shown in Fig. 20, we manually extract the center position of woven bag from the original 50 images with different illumination.

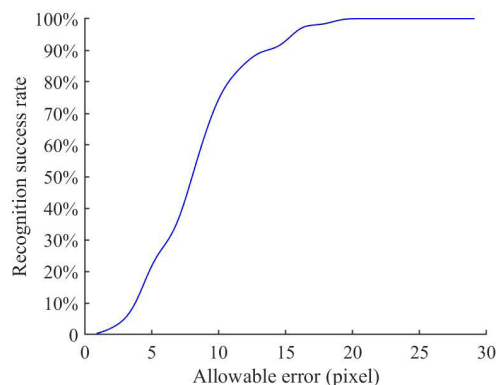




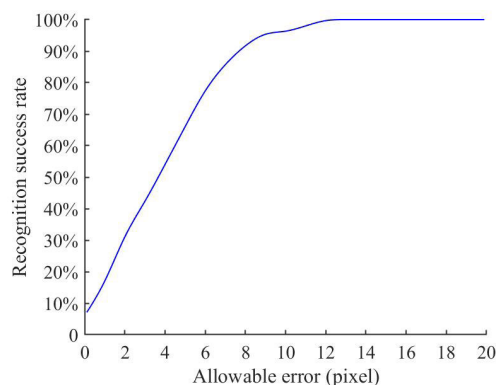
**FIGURE 20.** The center position of each woven bag by manually extraction.

In the actual robot de-palletizing working environment, the error is allowed in about 30 mm. For robotic de-palletizing tasks, before grabbing the woven bag, the image processing system is always moved in a horizontal guide, as shown in Fig. 1, which guarantees that the camera is always positioned at the same height, with the optical axis of the camera perpendicular to the ground. On the other hand, the height of woven bags is not always kept the same because stacking layers vary with time. However, the stacking layers are typical less than 7 to avoid collapsing risks and other potential dangers due to the collapsibility under great pressure of woven bags. When the stacking layers is only one, we measured the Euclidean diagonal distance of single woven bag in both image and real world. The results are 780.45 pixels and 778.53 mm, respectively, so the ratio between them is 0.9975 mm/pixel, approximately 1 mm/pixel. So 30mm in real world corresponds to around 30 pixels in the image. When the stacking layers increased, the height of upper layer is bigger, indicating the decreased ratio. When the stacking layers is up to five, this ratio is attenuated to approximately 0.8 mm/pixel. Under this circumstance, 30 mm in real world corresponds to around 37.5 pixels in the image. Hence the minimal error in pixels is set as low as 30 pixels in this paper to guarantee success for all circumstances.

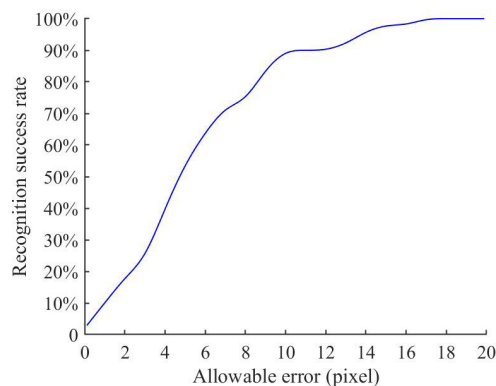
Fig. 21a,b,c are the relationships between the allowable error pixels and the recognition success rate of 50 woven bag images in the Euclidean distance and distance along the horizontal and vertical direction. From Fig. 21a, when the allowable error is less than 5 pixels in the Euclidean distance, the success rate is relatively low. As allowable error pixel increases, the success rate is continuously improved; when allowable error is greater than 13 pixels, the success rate can reach more than 90%; when allowable error reaches 19 pixels, the success rate achieves 100%. From Fig. 21b,c, the recognition success ratio reached 100% with allowable error along horizontal and vertical axis being 12 and 17 pixels. One possible explanation for the difference is, for six-flower vertical stack, the longer side of woven bags happens to be parallel to the vertical axis. The extracted center coordinate and orientation of the woven bag is obtained by straight line extraction using PPHT, and further fitting a rectangle. As is known to us, the woven bags are readily deformable, supposing the



(a)



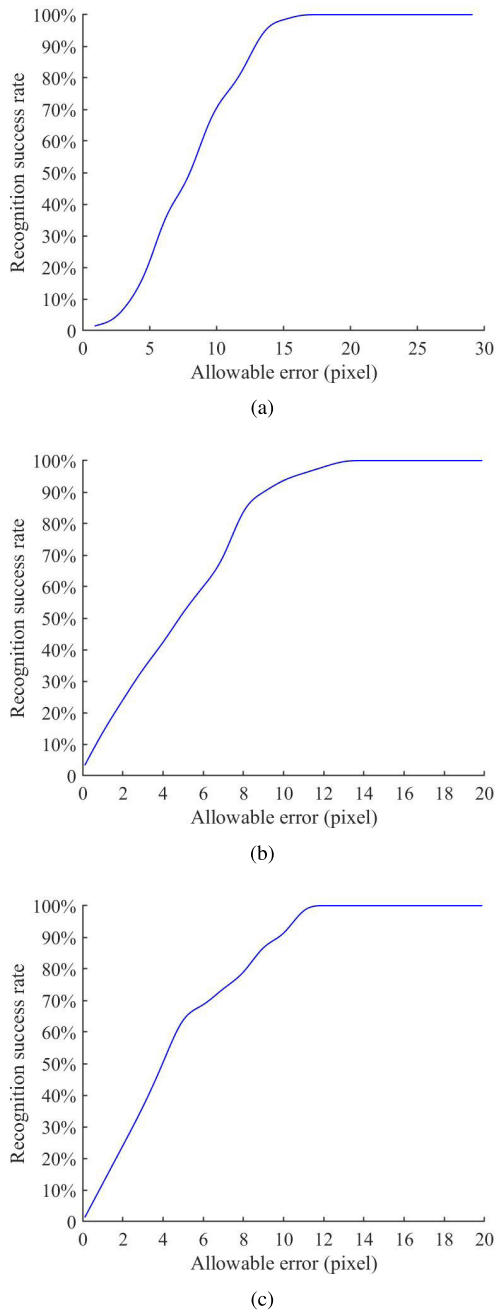
(b)



(c)

**FIGURE 21.** The relationship between the allowable error pixel values and the recognition accuracy. (a) Recognition success rate versus Euclidean distance. (b) Recognition success rate versus distance along the horizontal direction. (c) Recognition success rate versus distance along the vertical direction.

coordinate of each corner changed approximately the same along vertical and horizontal axis because of deformation, because the vertical sides of woven bags are longer, the horizontal coordinate of the extracted center point tends to be more accurate. Under such a circumstance, the allowable error along the horizontal axis is smaller compared to the one along the vertical axis when the success rate reached 100%.



**FIGURE 22.** The relationship between the allowable error pixel values and the recognition accuracy. (a) Recognition success rate versus Euclidean distance. (b) Recognition success rate versus distance along the horizontal direction. (c) Recognition success rate versus distance along the vertical direction.

Then we test the images of six-flower horizontal stack in the same way. The relationships between the allowable error pixels and the recognition success rate of 50 images in the Euclidean distance and distance along the horizontal and vertical direction are respectively shown in Fig. 22a,b,c.

From Fig. 22a, when the allowable error is less than 5 pixels in the Euclidean distance, the success rate is relatively low. As allowable error pixel increases, the success rate is

continuously improved; when allowable error is greater than 13 pixels, the success rate can reach more than 90%; when allowable error reaches 16 pixels, the success rate achieves 100%. From Fig. 22b,c, the recognition success ratio reached 100% with allowable error along horizontal and vertical axis being 13 and 11 pixels. Different from six-flower vertical stack, the longer side of woven bags happens to be parallel to the horizontal axis for six-flower horizontal stack. Under such a circumstance, the allowable error along the vertical axis is smaller compared to the one along the horizontal axis when the success rate reached 100%.

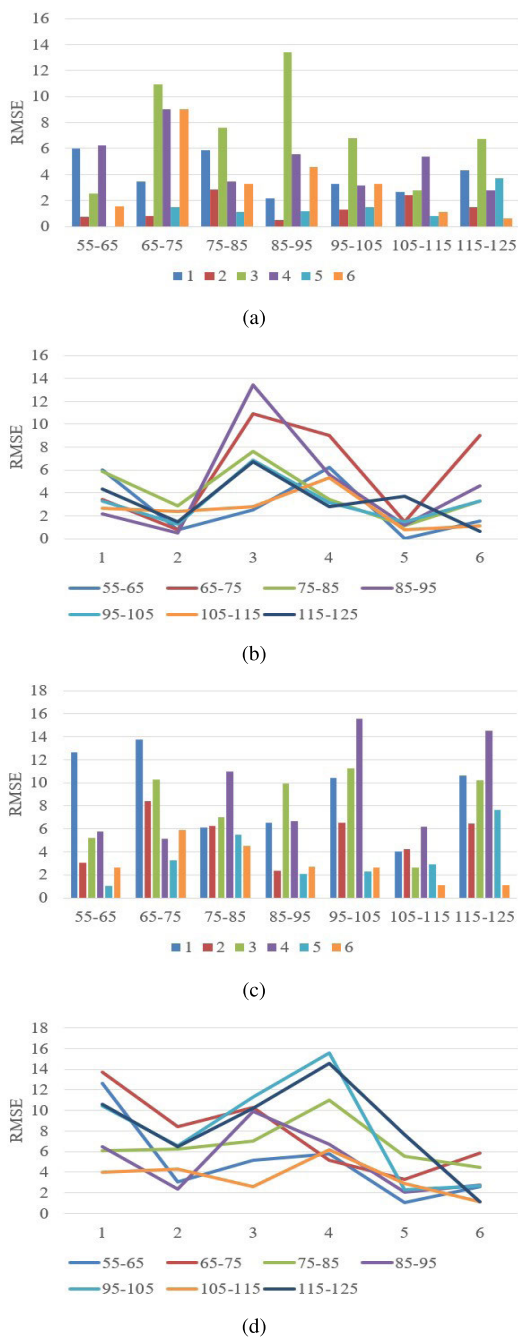
The results of the above experiments on different palletizing types all satisfy the requirement, proving the feasibility of the algorithm. And through error analysis, the identification error of the woven bag is small and the accuracy is high, proving the accuracy of the system to identify and locate the woven bag. The workpiece localization algorithm runs on a PC with an Intel(R) Core(TM) i5-8265U CPU (1.60 GHz) and 8 GB RAM. The time consuming for each layer's vision processing is 500ms on the average.

### C. THE ROBUSTNESS OF THE SYSTEM TO THE CHANGE OF ILLUMINATION

Considering the obstacle of variable ambient light on which we mainly focus, we test the robustness on illumination variation of the system. The original picture data of the six-flower vertical stack type are expanded, to calculate the Root Mean Square Error (RMSE) of each woven bag under different illumination intervals. We collect 50 images of six-flower horizontal stack with the same position, under different illumination. According to their brightness values, images are divided into seven illumination intervals (55-65, 65-75, 75-85, 85-95, 95-105, 105-115, 115-125). For the experiment, select 5 images from each interval as sample pictures, and then calculate the RMSE of each woven bag's coordinates in each interval. The results and analyses are as follows:

Fig. 23a,c RMSE-illumination-intervals graph representing histograms of RMSE of center positions'  $x$  and  $y$  coordinates respectively, in which, six different color of bars represent positions of each woven bag in six-flower vertical stack. The smaller the value of RMSE, the smaller the volatility of this set of coordinates under different light. From Fig. 23a,c, the most RMSEs of  $x$  coordinates are below 4 pixels, and  $y$  coordinates are below 6 pixels. The small RMSE value means the system can accurately identify and locate the woven bag in all illumination intervals, which also proves the robustness of the system to the change of illumination. And 105-115 is the optimal illumination interval with the best results, comparing with the overall RMSE of the other six illumination intervals.

As shown in Fig. 23b,d, RMSE-positions graph reflects the RMSE of six woven bags' positions in each image. The smaller the value of RMSE, the smaller the volatility of this set of coordinates in different positions. Comparing the overall RMSE of the six locations, we observed that the RMSEs of 2 and 5 are obviously smaller than other positions, because



**FIGURE 23.** the Root Mean Square Error (RMSE) of each woven bag’s coordinates in each interval. (a) RMSE-illumination-intervals graph representing histograms of RMSE of center positions’ x coordinates. (b)RMSE-positions graph reflects the RMSE of center positions’ x coordinates of six woven bags’ positions in each image. (c) RMSE-illumination-intervals graph representing histograms of RMSE of center positions’ y coordinates. (d)RMSE-positions graph reflects the RMSE of center positions’ y coordinates of six woven bags’ positions in each image.

the corresponding positions of 2 and 5 are both in the middle, with relatively small edge deformation.

On the whole, the main reasons generating failed recognition are presumably as follows. First, the intensity variation of image caused by the varied ambient light. Our algorithm

is based on the acquired image, which is readily affected by the unstable ambient light. Even if the influence of ambient light is considered in our method, its influence can hardly be utterly reduced. Second, foreground extraction error. At the proposed method, the first step is to extract the whole stack from the image. However, the extraction will be affected by the background, when the background is similar to the stack. Third, the characteristics of woven bags themselves. As repeatedly mentioned in this paper, the woven bags are prone to deform, which makes it difficult to ascertain its center and orientation. Moreover, the edge of different woven bags is similar to each other, which may cause recognition errors.

## VI. CONCLUSION

In this paper, we propose a two-step technique to accomplish the de-palletizing task under the environment of unstable natural light. Firstly, the whole stack is segmented from background interference. Secondly, based on the whole stack extracted, the optimized PPHT algorithm is used to identify and locate a single woven bag. Among them, aiming at the problems of illumination instability, low image contrast and difficult segmentation in the actual de-palletizing job, an effective Region Growing image segmentation method based on the dynamic threshold is proposed. This method establishes the relationship model between illumination and optimal segmentation threshold, and enhances the robustness of image segmentation to illumination change. Then, three optimal parameters of PPHT function are ascertained for different light intensity intervals, and a single woven bag is located by PPHT with optimal parameters. On this basis, the position and orientation information of each workpiece can be obtained from the minimal circumscribed rectangle, which can further guide the robot to complete the de-palletizing action. Experiments show the proposed technique is valid under stable natural light environment. As application-oriented engineers, we will further focus on challenges in workpiece localization, which include identification of heterogeneous surfaces, over-exposure induced by objects with high reflectivity, algorithm acceleration and optimization under limited computing power among others.

## REFERENCES

- [1] Y. Xu, Y. Liu, L. Hao, and H. Cheng, “Design of palletizing algorithm based on palletizing robot workstation,” in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Jun. 2016, pp. 609–613.
- [2] M. Pajor and M. Grudziński, “Intelligent machine tool—Vision based 3D scanning system for positioning of the workpiece,” *Solid State Phenomena*, vols. 220–221, pp. 497–503, Jan. 2015.
- [3] M. Kavoussanos and A. Pouliezios, “Visionary automation of sack handling and emptying,” *IEEE Robot. Autom. Mag.*, vol. 7, no. 4, pp. 44–49, Dec. 2000.
- [4] J. Norberto Pires, “Handling production changes online: Example using a robotic palletizing system for the automobile glass industry,” *Assem. Autom.*, vol. 24, no. 3, pp. 254–263, Sep. 2004.
- [5] G. ElMasry, S. Cubero, E. Moltó, and J. Blasco, “In-line sorting of irregular potatoes by using automated computer-based machine vision system,” *J. Food Eng.*, vol. 112, nos. 1–2, pp. 60–68, Sep. 2012.
- [6] N. Herakovic, *Robot Vision in Industrial Assembly and Quality Control Processes*. London, U.K.: Robot Vision. IntechOpen, 2010.

- [7] J. Li, Z. Chen, G. Rao, and J. Xu, "Structured light-based visual servoing for robotic pipe welding pose optimization," *IEEE Access*, vol. 7, pp. 138327–138340, 2019.
- [8] J. Li, Z. Chen, and J. Xu, "Fringe projection based visual servoing for cylindrical surface positioning task," *Proc. SPIE*, vol. 11205, Oct. 2019, Art. no. 112051N.
- [9] J. Y. Chang and J. Y. Shen, "Machine vision assistance for flex cable-actuator comb manufacturing in hard disk drives," in *Proc. 15th Int. Conf. Mechatronics Mach. Vis. Pract.*, Dec. 2008, pp. 195–200.
- [10] W. Huang and R. Kovacevic, "Development of a real-time laser-based machine vision system to monitor and control welding processes," *Int. J. Adv. Manuf. Technol.*, vol. 63, nos. 1–4, pp. 235–248, Nov. 2012.
- [11] R. T. Chin and C. A. Harlow, "Automated visual inspection: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 6, pp. 557–573, Nov. 1982.
- [12] T. Yuan-yuan, L. Si-yang, and T. Qing-chang, "Application of detecting Part's size online based on machine vision," *Energy Procedia*, vol. 16, pp. 1948–1956, Jan. 2012.
- [13] K. Rahardja and A. Kosaka, "Vision-based bin-picking: Recognition and localization of multiple complex objects using simple visual cues," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. IROS*, Nov. 1996, pp. 1448–1457.
- [14] J. Guo, X. Liu, and X. Chen, "Automatic measurement of the mechanical part based on binocular stereo vision," in *Proc. 7th World Congr. Intell. Control Autom.*, Jun. 2008, pp. 8487–8491.
- [15] X. Fan, X. Liu, X. Wang, and Y. Xiao, "An automatic robot unstacking system based on binocular stereo vision," in *Proc. IEEE Int. Conf. Secur., Pattern Anal., Cybern. (SPAC)*, Oct. 2014, pp. 86–90.
- [16] W.-Y. Chang, J.-W. Hsu, and B.-Y. Hsu, "3D scanning system of structured light for aiding workpiece position of CNC machine tool," in *Proc. IEEE Int. Conf. Adv. Manuf. (ICAM)*, Nov. 2018, pp. 388–391.
- [17] R. Dieter Schraft and T. Ledermann, "Intelligent picking of chaotically stored objects," *Assem. Autom.*, vol. 23, no. 1, pp. 38–42, Mar. 2003.
- [18] J. K. Oh and C. H. Lee, "Development of a stereo vision based position measurement algorithm for the randomly piled objects," in *Proc. Int. Conf. Adv. Robot.*, Aug. 2007, pp. 64–69.
- [19] B. Zhang and S. B. Skaar, "Robotic de-palletizing using uncalibrated vision and 3D laser-assisted image analysis," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2009, pp. 3820–3825.
- [20] S. J. Eun, E. Y. Jung, D. K. Park, and J. Y. Kim, "Effective fat quantification using multiple region growing scheme at high-field MRI," *IEEE Access*, vol. 7, pp. 2118–2125, 2019.
- [21] Y. Xu, Z. Mao, C. Liu, and B. Wang, "Pulmonary vessel segmentation via stage-wise convolutional networks with orientation-based region growing optimization," *IEEE Access*, vol. 6, pp. 71296–71305, 2018.
- [22] C. Zhang, L. Wangsheng, and J. Xie, "Based on CCD imaging ball outside the new algorithm for extracting profile," in *Proc. Int. Conf. Electr. Control Eng.*, Jun. 2010, pp. 529–531.

• • •