Three-dimensional glue detection and evaluation based on linear structured light and Multi-task Cascaded CNN

Lei Geng [1, 2, a], Ruipeng Yang [1, 2, b], Zhitao Xiao [1, 2, c, *] and Yanbei Liu [1, 2, d]

[1]Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tianjin, 300387, China;

[2]School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin, 300387, China.

[a]genglei@tjpu.edu.cn, [b]654190545@qq.com, [c]xiaozhitao@tjpu.edu.cn,

[d]liuyanbei@tjpu.edu.cn

Abstract

The quality of the gluing detection directly affects the technical level of the robot gluing. Glue detection method based on traditional machine vision can not meet the demands of precision and real-time of high-speed measurement system in the case of uneven illumination. A 3D glue detection method based on the linear structured light and the Multi-Task Cascaded CNN is proposed. Firstly, a cascaded structure with three stages of deep convolutional networks is implemented to coarsely locate the region and key points of the glue. Then, a multi-objective optimization model is established to locate the key points accurately which is based on the key point coordinates, the gray center and the epipolar constraint. Finally, the 3D coordinates of the key points, the height and width of the glue are calculated. Experiments show that the absolute error of measurement is less than 0.99mm and the relative error is less than 9.7%. The proposed method can obviously improve the detection performance and reduce the error detection.

Keywords

Glue detection, binocular vision, Multi-Task Cascaded CNN, linear structured light.

1. Introduction

As gluing is more and more widely used in high-precision devices, the accuracy of gluing becomes a highlighted problem. In the process of gluing for precision equipment, the glue has more stringent specifications in the amount of glue, the location of glue, the width of glue and mixing ratio. Traditional gluing robots are weak in the process of automatic inspection for the quality of gluing. At present, the purpose of traditional 2D gluing detection is the localization and segmentation of glue, which can not estimate the uniformity of glue with complex shape[1,2]. Therefore, the 2D detection of the glue can not meet the requirements of product quality testing in some industry; on the contrary, the 3D detection of the glue can improve product quality and decrease scrap rate, so it has important research value.

There are still many problems in the study of 3D glue detection. The three-dimensional recovery system based on the structured light has the advantages of high precision and strong anti-interference ability. Therefore, The accuracy of gluing detection and the time cost of detection are the key points and research focus in 3D gluing detection[3]. Mao et al.[4] puts forward a binocular vision measurement system based on line structure light to realize 3D volume measurement. Epipolar constraint is implemented to match the homonymy points and restore the 3D point cloud of the

structure light strip. However, the measurement is redundant because of the 3D recovery of the whole structured light stripe. Lin et al.[5] puts forward a passive line structured light scanning measurement system to obtain 3D data of engine cylinder. The system uses one camera to recover the structured light, so repeated measurement extends the error of the camera carrier. Ma et al.[6] puts forward a 3D surface reconstruction method using a binocular stereo vision technology and a coded structured light, which combines a gray code with phase-shift has been studied. This method has a lot of redundant information too, so the measurement is inefficient.

It can be seen from the previous text, the traditional 3D recovery methods based on the structured light often need to restore the 3D point cloud of the whole structure light projected area. The really useful parts are not so much, and it is more complicated to look for objects in 3D point cloud than a 2D image. So, if we can locate the object which will be measured before the 3D restoration, we can do 3D recovery purposefully and reduce redundant calculations. In this paper, a Multi-Task Cascaded CNN is implemented to detect the key points of the structured light strip and we use a multi-objective optimization model to improve the accuracy of the system. So, we reduce computational redundancy while increasing the accuracy of the system and improve the system efficiency.

2. System Model

The measurement steps mainly include system calibration, coarsely locating of the key points and the glue region in structured light strip, accurately locating of the key points, key points restoration and evaluating the quality of the glue. The part of system calibration includes the camera internal calibration, camera external calibration and robot hand-eye calibration. It transfers the points from two-dimensional image coordinates to the world coordinate. A cascaded structure with three stages of deep convolutional networks is implemented to coarsely locate the region and key points of the glue. A multi-objective optimization model is established to locate the key points accurately by iteratively calculating the coordinates of key points in left and right images. The part of 3D point cloud restoration firstly solves the 3D point cloud coordinates. Then it unified the results of multiple measurements into the same coordinate system by the results of hand-eye calibration. Finally, the width, height, and continuity of glue will be calculated. System structure is shown in Figure 1. The line structured light and binocular camera are fixed at the Tool Center Point(TCP) of the robot by the customized connector. The structured light is vertically irradiated on the glue and the binocular camera acquire image from the side[7-9].

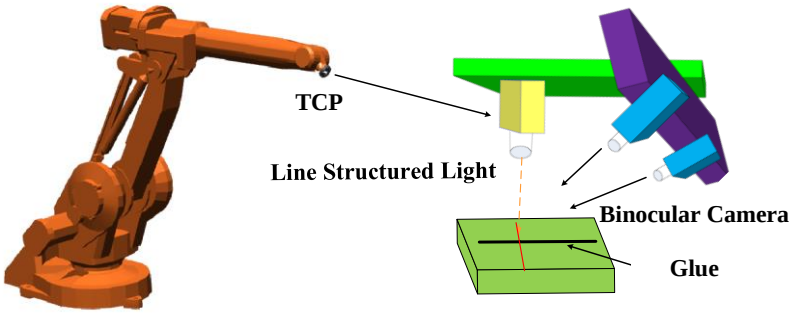

Fig. 1 The structure of the proposed system

System flow chart shown in Fig. 2, it includes computing the intrinsic parameters of camera, binocular calibration, hand-eye calibration, coarsely locating of the key points and the glue region in structured light strip, accurately locating of the key points, key points restoration and evaluating the quality of the glue.

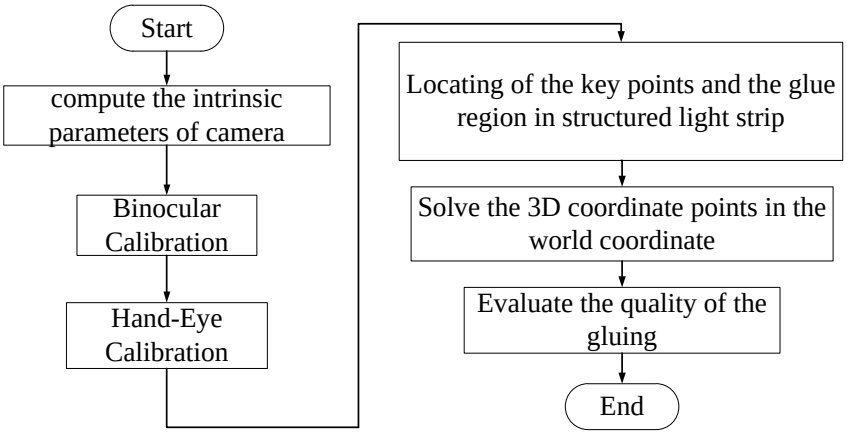

Fig. 2. The flowchart of the proposed method

3. Method

3.1 System Calibration

The part of system calibration includes camera calibration and robot hand-eye calibration. Camera calibration includes the internal calibration of two cameras and the calibration of the relative position between the two cameras (Translation vector T and rotation matrix R from right camera to left camera). In this paper, Zhang's calibration method is used. The method requires to take multiple images of the plane template from different angles. Through the correspondence between each feature point on the plane template and the point on its image, we optimize the results of the calculation. Finally, we use the internal parameter matrix and the homography matrix to find the corresponding external parameters [10,11].

3.2 Coarsely Locating of the Key Points and the Glue Region

In the process of 3D reconstruction of structured light strips, the structured light strips are purposefully restored if the gluing regions are located. The implementation of this process can reduce the extra work of 3D restoration greatly and reduce the time cost. In order to improve the efficiency of 3D point measurement, we use a multi-task cascade CNN [12]to locate the gluing regions and the key points in the structured light strips. The network is a three-layer convolutional neural network, including P-Net, R-Net and O-Net. The introduction and network structure of each layer are as follows.

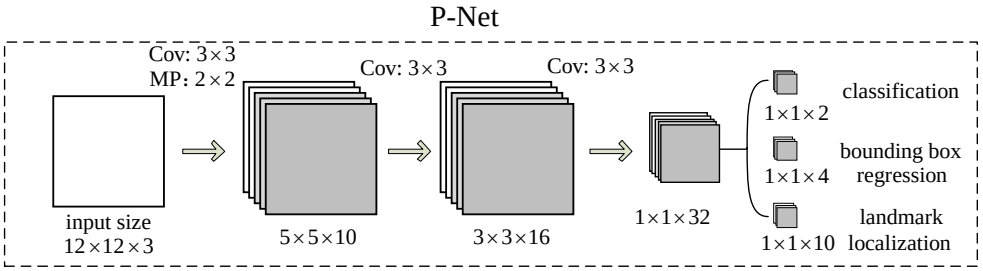

Fig. 3. The architectures of P-Net

The structure of the P-Net is shown in Figure 3, where “MP” denotes max pooling and “Conv” denotes convolution, the stride of convolution and pooling is 1 and 2, respectively. This network can be used to obtain the regression coefficient vector of candidate windows and bounding boxes of the gluing area. In detail, the bounding boxes are regressed, the candidate windows are calibrated, and the overlapping candidate windows are merged by non-maximal suppression (NMS) [13].

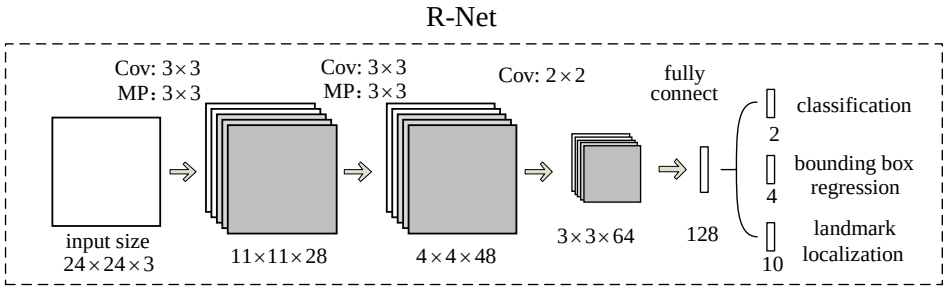

Fig. 4. The architectures of R-Net

The structure of the R-Net is shown in Figure 4, which is used to refine the results of the P-Net network. Through this network, we can get finer candidate area. In the step above, the false-positive regions is removed by the bounding box regression and NMS.

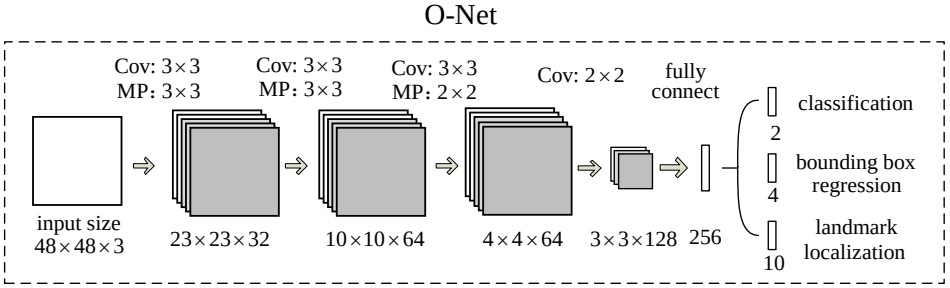

Fig. 5. The architectures of O-Net

The structure of the O-net is shown in Figure 5. This network has one more layer of convolution over R-Net network, so its results will be finer than R-Net. The function of this network is same to the R-Net network. However, the network has more conditions for the detection area. It could output 5 landmarks in the end.

The above network was used in the gluing region detection of structured light strips and coarsely locating of key points, as shown in Figure 6. It could locate five key points, including the glue apex point, the left endpoint, the right endpoint, the left bottom endpoint and the right bottom endpoint. Finally, we calculate the width and height of the glue by restoring the 3D coordinates of the five key points.

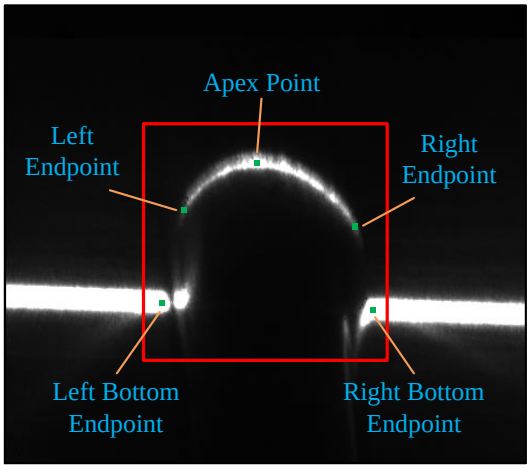

Fig. 6. Locating of the key points and the glue region

3.3 Accurately Locating of the Key Points

For the primary locating of the key points of glue, coordinate accuracy can not be accurate to sub-pixel accuracy and the key points of left and right image may not be on the same epipolar plane. We proposed a method to locate the key points of left and right image, correct the coordinates to the same epipolar plane. Firstly, a multi-objective optimization is implemented to correct a key point. Then we

update the coordinates of the key point and iteratively calculate the key point of the left and right images[14,15].

3.3.1 Key Point Optimization Method

Assuming the corrected key point coordinates is $K(x,y)$. A objective function of multi-objective optimization model was propose.

$$\text{Min } D_{k-k_0} 、 D_{k-l} 、 D_{k-k_c} \tag{1}$$

$$\begin{cases} 0 < x < I_{width} \\ 0 < y < I_{Height} \\ |x - x_0| < H_l \\ |y - y_0| < H_l \end{cases} \tag{2}$$

As shown in formula (1), D_{k-k_0} represents the distance between the point K and the key point, D_{k-l} represents the distance between the point K and the epipolar line which is calculated by the corresponding key point in the other image, D_{k-k_c} represents the distance between the point K and the grayscale center in the $3\times H_l$ pixel block centered on the key points. The objective function of multi-objective optimization is established to minimize D_{k-k_0}, D_{k-l} and D_{k-k_c}. So, we calculate the optimal result under the maximum & minimization model by formula 2. I_{width} and I_{Height} represent the width and height of the image respectively.

3.3.2 Iterative Optimization Method

The precise position of the key point is obtained by solving the multi-objective optimization problem. Then the key points are replaced by computed results. So, the grayscale center and the epipolar line calculated by the key point also changed. After calculating the coordinates of the key point, we can continue iteratively calculating the new key points in the other image. The iteration can be stopped until all distances are less than the reprojection error of the system calibration.

3.4 Binocular Matching and 3D Restoration

3.4.1 Binocular matching

As shown in Figure 7, the points P_1 and P_2 are the corresponding points of the spatial point P in space. The image points P_1 and P_2 on the two cameras C_1 and C_2 have been separately detected from the two images. The projection matrices of cameras C_1 and C_2 are M_1 and M_2 respectively. So, we can get:

$$Z_{c1}\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^1 & m_{12}^1 & m_{13}^1 & m_{14}^1 \\ m_{21}^1 & m_{22}^1 & m_{23}^1 & m_{24}^1 \\ m_{31}^1 & m_{32}^1 & m_{33}^1 & m_{34}^1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3}$$

$$Z_{c2}\begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^2 & m_{12}^2 & m_{13}^2 & m_{14}^2 \\ m_{21}^2 & m_{22}^2 & m_{23}^2 & m_{24}^2 \\ m_{31}^2 & m_{32}^2 & m_{33}^2 & m_{34}^2 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4}$$

Among them, $(u_1, v_1, 1)$ and $(u_2, v_2, 1)$ are the image homogeneous coordinates of the points P_1 and P_2 in the respective images; $(X, Y, Z, 1)$ is the homogeneous coordinates of point P in the world

coordinate system; m_{ij}^{k} is an element of m^{k}, located in row i and column j. The formula (3) and (4) obtain the four linear equations for X, Y, and Z by eliminating terms of Z_{c1} and Z_{c2}.

$$\begin{cases} \left(u_1 m_{31}^1 - m_{11}^1\right) X + \left(u_1 m_{32}^1 - m_{12}^1\right) Y + \left(u_1 m_{33}^1 - m_{13}^1\right) Z = m_{14}^1 - u_1 m_{34}^1 \\ \left(v_1 m_{31}^1 - m_{21}^1\right) X + \left(v_1 m_{32}^1 - m_{22}^1\right) Y + \left(v_1 m_{33}^1 - m_{23}^1\right) Z = m_{24}^1 - v_1 m_{34}^1 \\ \left(u_2 m_{31}^2 - m_{11}^2\right) X + \left(u_2 m_{32}^2 - m_{12}^2\right) Y + \left(u_2 m_{33}^2 - m_{13}^2\right) Z = m_{14}^2 - u_2 m_{34}^2 \\ \left(v_2 m_{31}^2 - m_{21}^2\right) X + \left(v_2 m_{32}^2 - m_{22}^2\right) Y + \left(v_2 m_{33}^2 - m_{23}^2\right) Z = m_{24}^2 - v_2 m_{34}^2 \end{cases} \tag{5}$$

Since the spatial point P is the intersection of line O_1P_1 and line O_2P_2, it must satisfy both the constraints represented by the formula (5). We combine these four equations to find the coordinates of point P. Because we have assumed P_1 and P_2 are the corresponding points of the spatial point P on different views, line O_1P_1 and O_2P_2 must intersect. The homogeneous coordinates of point P could be obtained by least squares method.

3.4.2 Evaluation of the Quality of Gluing

The part of gluing quality evaluation mainly completes the calculation for the width and height of the glue. The methods of calculation are shown in Figure 8. The point A, B, C, D and E represent the glue apex point, the left endpoint, the right endpoint, the left bottom endpoint and the right bottom endpoint., respectively. M_{bd} and M_{ce} represent the midpoint of the line connecting B, D and the midpoint of the line connecting C, E, respectively. The distance between the point A and the line connected by the point C to the point D is the height of the glue. The distance between the point M_{bd} and the point M_{ce} is the width of the glue.

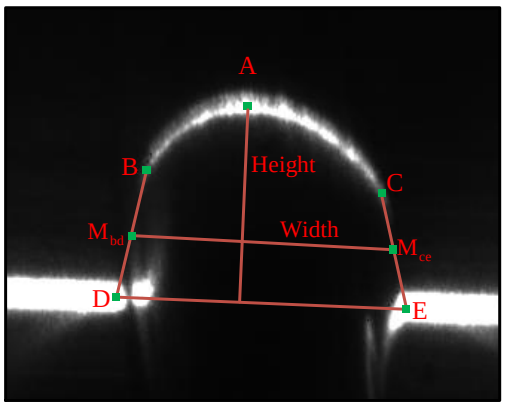

Fig. 7. The evaluation method of the gluing quality

4. Experiment and Analysis

4.1 Data preparation and model training

The neural network was trained with 1009 images. In order to expand the data, the training of the neural network is performed on patches of the preprocessed full images, the patches shown in Figure 8. The training data is much larger than the number of training images. Each patch is obtained by randomly selecting its center inside the full image. There are 150000 patches obtained by randomly extracting in each layer. The first 90% of the dataset is used for training, while the last 10% is used for validation. We use four different kinds of data annotation in our training process: (i) Negatives: Regions that the Intersection-over-Union (IoU) ratio less than 0.3 to any ground-truth; (ii) Positives: IoU above 0.65 to a ground truth; (iii) Part faces: IoU between 0.4 and 0.65 to a ground truth; and (iv) Landmark: labeled 5 landmarks' positions. Negatives and positives are used for classification tasks,

positives and part data are used for bounding box regression, and landmark data are used for landmark localization[12].

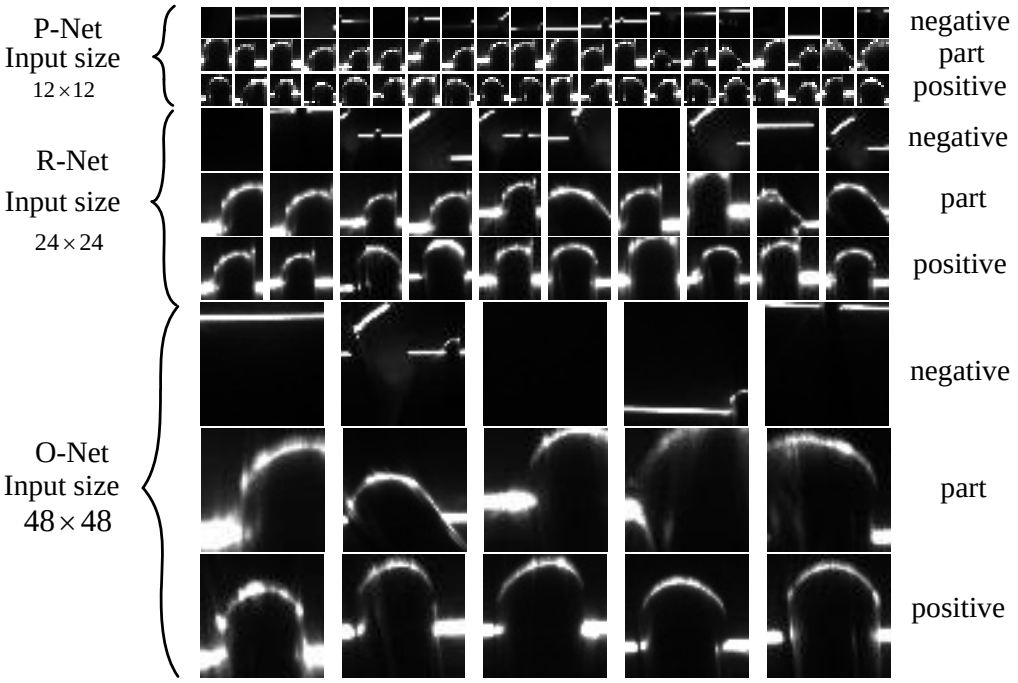

Fig. 8 Patches of the preprocessed full images

The input images come with their corresponding fully annotated ground truth are used to train the network with the stochastic gradient descent implementation of Caffe. Accordingly we use a high momentum (0.9) such that a large number of the previously seen training samples determine the update in the current optimization step[16]. The batchsize and base learning rate are 64 and 0.001 respectively. Thanks to data augmentation with extracting patches from images, it only needs few annotated images and has a reasonable training time of 8 hours on a NVIDIA GTX 1080Ti GPU (11 GB). The Accuracy-Loss curve of training process is shown in Fig. 9.

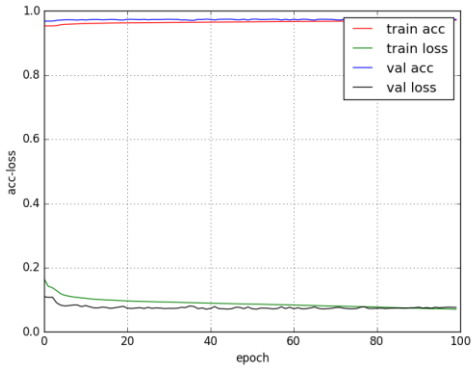

Fig. 9 Acc-Loss curve

4.2 Evaluation Criteria and Measurement Results.

The length and height of the test glue are measured by HEXAGON ROMER73 which has higher measurement accuracy. We calculate the mean absolute error (MAE), mean relative error (MRE) and mean time (MT) to measure the precision of the three-dimensional gluing detection system.

In the experiment, three kinds of glue were compared including white glue, black glue and blue glue and we repeat the experiment with Mao et al. [4], Lin et al. [5], Ma et al. [6] and our method. Result is as shown in Table 1, Table 2 and Table 3.

Table 1. The comparison of the white glue detection

	MAE(mm)	MRE(%)	MT(ms)
Mao et al. [4]	2.27	12.6	251.3
Lin et al. [5]	2.01	13.0	201.3
Ma et al. [6]	1.12	9.9	318.0
Our method	**0.67**	**6.7**	**106.6**

Table 2. The comparison of the blue glue detection

	MAE(mm)	MRE(%)	MT(ms)
Mao et al. [4]	2.44	21.1	196.5
Lin et al. [5]	3.23	10.1	238.7
Ma et al. [6]	1.19	37.9	317.9
Our method	**0.93**	**8.7**	**116.5**

Table 3. The comparison of the black glue detection

	MAE(mm)	MRE (%)	MT(ms)
Mao et al. [4]	3.38	31.9	206.5
Lin et al. [5]	4.35	32.3	231.6
Ma et al. [6]	2.21	19.1	313.6
Our method	**0.99**	**9.7**	**108.9**

We compare our schedule with another three methods in detection of three kinds of glue in Table 1, Table 2 and Table 3. It is found that it has almost indifferent performance among every method because of the strong light reflection of the structured light by glue, such as white glue and blue glue. Our method have the less error with the other methods. While the absolute error and relative error of our method less than another three methods for black glue which has stronger ability to absorb the structured light. And our method has a corresponding shortening in the average time.

One of the advantages of our approach is that we use a multi-task cascade CNN, so we eliminate the redundancy of the structured light strip and the complexity of the algorithm. At the same time, a multi-objective optimization model based on cyclic iteration is proposed to accurately locate the key points of the structured light, so the accuracy of the detection system is inproved. It can be seen that this method can significantly improve the accuracy and speed of glue detection and we can also measure the 3D information of the glue by the proposed system.

5. Conclusion

In the robot glue detection, the traditional method of three-dimensional gluing detection can not meet the requirement of precision and real time. A new glue detection solution based on linear structured light and Multi-task Cascaded CNN is proposed in this paper. The quality evaluation of the gluing was realized by system calibration, coarsely locating of the key points and the glue region in structured light strip, accurately locating of the key points, key points restoration and evaluating the quality of the glue. Experiments show that the absolute error of measurement is less than 0.99mm and the relative error is less than 9.7% respectively. It can satisfy the real-time performance of the system and has strongly application value.

Acknowledgements

This work was supported by National Natural Science Foundation of China under grant No. 61771340, Tianjin Science and Technology Major Projects and Engineering under grant No. 17ZXSCSY00090, Basic Application Research Project of China National Textile and Apparel Council under grant No.

J201509, Major Program of National Natural Science Foundation of China under grant No. 13&ZD162, Plan Program of Tianjin Educational Science and Research under grant No.2017KJ087.

References

[1] Wang N, Liu J, Wei S, et al. A Vision Location System Design of Glue Dispensing Robot [C]. International Conference on Intelligent Robotics and Applications, 2015: 536-551.

[2] Zhang J, Yang J, Li B. Online Detection and Control System for Auto Body Welding Fixture [C]. International Conference on Measuring Technology and Mechatronics Automation. IEEE, 2009: 816-819.

[3] Tang D W, Zong D X, Deng Z Q, et al. On Application of Glue-robot Vision System [J]. Robot, 2006, 28(1): 1-4. (in Chinese)

[4] Mao J, Lou X, Li W, et al. Binocular 3D volume measurement system based on line-structured light [J]. Optical Technique, 2016, 42(1). (in Chinese)

[5] Lin Z, Wang T, Nan G, et al. Three-dimensional data measurement of engine cylinder head blank based on line structured light [J]. Opto-Electronic Engineering, 2014(5): 46-51. (in Chinese)

[6] MA Z, Han F, Wang T. 3D surface reconstruction based on binocular vision using structured light [J]. Journal of Beijing Institute of Technology (English Edition), 2016, 25(3): 413-417.

[7] Wang Y. Research on online inspection technology for robot glue spreading quality based on machine vision [D]. Harbin: Harbin Institute of Technology, 2015. (in Chinese)

[8] Wang P. Study on Key Techniques for Automatic 3D Structured-Light Scanning System [D]. Tianjin: Tianjin University, 2008. (in Chinese)

[9] Dornaika F, Horaud R. Simultaneous robot-world and hand-eye calibration [J]. IEEE Transactions on Robotics and Automation, 2011, 14(4): 617-622.

[10] Zhang Z Y. A Flexible New Technique for Camera Calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.

[11] Moldovan D, Wada T. A calibrated pinhole camera model for single viewpoint omnidirectional imaging systems [C]. IEEE International Conference on Image Processing, Singapore, 2004: 2977-2980.

[12] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks [J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

[13] Kehu X U, Wang T, Chen J. Harris Corner Detection Algorithm Based on Self-adapting Non-maximal Suppression [J]. Science & Technology Review, 2013, 73(3): 931-947.

[14] Averbakh I, Lebedev V. Interval data minmax regret network optimization problems[J]. Discrete Applied Mathematics, 2004, 138(3): 289-301.

[15] Fonseca C M, Fleming P J. An Overview of Evolutionary Algorithms in Multiobjective Optimization [J]. Evolutionary Computation, 2014, 3(1): 1-16.

[16] Noh H, Hong S, Han B. Learning Deconvolution Network for Semantic Segmentation [J]. 2015: 1520-1528.