Research Article

Research on the Industrial Robot Grasping Method Based on Multisensor Data Fusion and Binocular Vision

Shangyu Xie

New York University Tandon School of Engineering, New York University, Brooklyn, New York, NY 11201, USA

Correspondence should be addressed to Shangyu Xie; sx2109@nyu.edu

Received 17 March 2022; Revised 26 April 2022; Accepted 3 May 2022; Published 25 May 2022

1. Introduction

Most of the handling-type industrial robots working on production lines nowadays are operated by means of demonstration or preprogramming, which makes the production lines less flexible and cannot meet the flexible production requirements of the material handling system. As one of the most typical types of digital control equipment with a wide range of applications and high technical added value, robots are playing an increasingly important role in modern advanced production and manufacturing, and the development of robotics will be a strong driver for future production and social development [1].

While industrial robots were originally developed to free workers from the monotony of repetitive work and hazardous environments, in recent years, they have been introduced in factories and companies with the primary aim of improving productivity and quality [2].

The price of robots has fallen considerably compared to the past and will continue to fall in the future, but the cost of labour has risen worldwide and is high in individual countries and regions, which provides a wide market for industrial robots, and sales of industrial robots are growing at a fast rate [3]. There are two main applications of industrial robots in production: robotic work cells and robotic work production lines, and in foreign countries, robotic work production lines have become the main application of industrial robots [4]. The robot as the core of the automated production line to adapt to the modern manufacturing industry more varieties, less batch of flexible production development direction, has a broad market development prospects and strong vitality and has developed a variety of for the automotive, electrical machinery, and other industries of automated complete sets of equipment and production line products [5].

Robotics involves the cross-fertilisation of multiple disciplines, involving mechanics, electronics, computers, communications, control, and many other aspects. In the modern manufacturing industry, along with the expansion of industrial robots and the development of robotics, the degree of automation, intelligence, and networking of robots become higher and higher, and the functions that can be achieved are becoming more and more functional and better [6]. The connotation of robotics has changed to "an intelligent system with real action functions that flexibly applies robotics." [7, 8]. The communication between the upper and lower robots was realized via Ethernet; the grasping scheme of the hand claw
was designed for the specific target object to be grasped in the experiment; the control software was developed using both schemes. Vision control is performed by operating specific keys on the control panel, and vision control is performed by executing a macrovariable program, which finally achieves the fusion of multisensing data and binocular vision.

2. Related Work

In the study by Hu et al. [9], along with the progress of science and technology and the development of production technology, the research of industrial robots based on machine vision technology has received more and more widespread attention, and a large number of scientific and technological achievements have emerged and have been well applied in actual production. A lot of work has been done on the research of industrial robots based on machine vision, and there have been many breakthrough results. The arc welding robot seam tracking control system studied by [10] uses vision sensors and image processing with neural networks to obtain seam shape data for effective seam tracking.

In the study by Li et al. [11], the XR22, a parallel robot developed by [12] incorporating a vision system for picking up parts on high-speed conveyors, can quickly pick up random goods from the conveyor and place them in order in the box by 2D positioning. The latest “second generation” picking robot from ABB of Sweden, the FlexPicker IRB360, offers the advantages of a larger payload, faster operation, and smaller footprint and with the help of simple and effective 2D vision, can quickly pick up items on a conveyor at speed of up to 2 times/second [13]. The robot vision system developed by the Finnish company RTS is successfully used in the production of naval propeller propellers, where the propeller blades are precisely ground according to CAD models by means of machine vision monitoring [14].

3. Introduction to the Principle of 3D Vision Measurement

In order to be able to accurately reconstruct the 3D point cloud data of the visible surface of the scattered stacked workpieces in the material frame, this study adopts binocular stereo vision plus structured light gray code [15] and phase displacement [16] for the acquisition of 3D point cloud data. The basic principle of binocular ranging [17] is shown in Figure 1, and based on the similar triangle in Figure 1, the equation can be obtained: the distances from $p$ and $p'$ to the left edge of the respective imaging planes are $X_R$ and $X_T$, respectively, and $B$ is the length of the baseline between the left and right cameras.

$$\frac{B}{Z} = \frac{(B + x_1) - x_1}{Z - f}$$

$$Z = \frac{B \times f}{x_r - x_t} = \frac{B \times f}{d}$$

In Figure 1, $O_R$ and $O_T$ are the main light points of the left and right cameras, respectively. If $P$ is a target point to be measured in 3D space, the imaging points of $P$ in the imaging planes of the two cameras are $p$ and $p'$, the distances from $p$ and $p'$ to the left edge of the respective imaging planes are $X_R$ and $X_T$, respectively, $B$ is the length of the baseline between the left and right cameras, and $Z$ is the distance from point $P$ to the baseline, which is the spatial position of the point obtained from binocular 3D vision measurement the spatial position.

Figure 2 shows the hardware composition of the 3D vision measurement system, including two industrial cameras, a digital laser projection surface structure light projector, a six-axis robot, and a frame for storing workpieces. The process of 3D measurement of the visual surface of the workpieces in a scattered pile is to first project a number of gray code images and four phase-shift images onto the surface of the workpieces in a scattered pile using a DLP projector. At the same time, a camera is used to capture the images of the visible surface of the workpiece, as shown in Figure 3, so that the correspondence between the points $(p, p')$ on the left and right camera images can be determined according to the rules of gray code encoding, phase-shift, and the principle of polar line constraint of stereovision.

Based on the binocular measurement principle described in the literature [18], the 3D point cloud data of numerous workpiece surfaces in the field of view are calculated. The 3D point cloud reconstructed from the measurement of a scattered pile of workpieces is shown in Figure 4.

4. Target Workpiece Segmentation and 3D Matching Methods

4.1. 3D Matching with Model Point Cloud Generation. To determine the position of a workpiece in a loose pile, a template point cloud $\Omega_0$ must be obtained for 3D matching, and then, $\Omega_0$ must be matched in 3D with the measured and segmented 3D reconstruction point cloud $\Omega_t$ of the target workpiece in the frame to obtain the position and attitude of the target workpiece in the loose pile in the frame.
There are two ways to obtain a 3D matching template point cloud of a workpiece. Figure 5 shows a CAD theoretical model of the workpiece obtained from the production unit, which is discretely processed to obtain a point cloud of the entire external surface of the workpiece. Figure 6 shows a template point cloud formed by measuring, reconstructing, and stitching a single workpiece in situ using the 3D measurement system developed in this study. Although the template point cloud $\Omega_0$ generated by the CAD model has better detail, the method of obtaining the template 3D point cloud $\Omega_0$ by in situ measurement has the advantage of being simple to use and easy to operate. This study shows that the accuracy and effectiveness of the two methods in 3D matching are basically the same.

4.2. Target Workpiece Segmentation Algorithm. There is no good method for segmenting the target workpiece from a 3D point cloud of a scattered pile of workpieces. Considering that this measurement system uses binocular vision and reconstruction through multiple sequential images and that the projecto’s light projection pattern is controllable, it is possible to obtain a 2D image of the workpiece in a scattered pile by projecting a uniform, intensity-controlled illumination source. The two-dimensional image is then divided into different regions by segmentation of this two-dimensional image; then, the three-dimensional point clouds corresponding to the two-dimensional image regions are also divided differently in the three-dimensional scene, so that the three-dimensional point cloud data subsets belonging to different individual workpieces can be roughly distinguished. Segmenting the scattered workpieces in the
Common 2D image segmentation algorithms include the binarisation method, watershed method [19], and edge method [20]. Figure 6 shows the results of the test image segmentation process using the edge method. In this study, this method is used to segment the scattered piles of workpieces in the material frame on 2D images to initially identify the target workpieces and the 3D point cloud subsets to which they belong.
4.3. Three-Dimensional Matching Algorithm. When using ICP algorithms, it is often necessary to use the PCA (principal component analysis) method in order to obtain a good initial pose and thus improve the speed and accuracy of ICP alignment.

As shown in Figure 7, if the point cloud consists of \( m \) points, let \( X \) be a \( 3 \times m \) matrix containing the coordinates of all the points; the coordinates of the centre of gravity of the point cloud are \( \mu \). First, the point cloud is translated, and the origin of the new coordinate system is the centre of gravity of the point cloud \( \mu \). The new \( 3 \times m \) matrix of the point cloud is obtained after the translation transformation \( X_{\mu} \). By constructing the covariance matrix \( \text{cov} \ X \) and obtaining the eigenvectors corresponding to the largest eigenvalues of \( \text{cov} \)
$X$, we further construct the rotation transformation matrix $R$, which is orthogonal and satisfies the right-hand system rule; using the rotation transformation $R$ to rotate $X_\mu$, we obtain $X_{out}$.

$$X_\mu = X - \mu,$$

$$\text{cov} X = X_\mu \times X_\mu^T,$$  \hspace{1cm} (2)

$$X_{out} = R \times X_\mu.$$  

The template point cloud $\Omega_0$ and target point cloud $\Omega_1$ are both transformed by PCA (at this point, the template point cloud $\Omega_0$ and the target point cloud $\Omega_1$ are both described in their respective centre of gravity coordinate systems) and then further matched by ICP to obtain the final 3D point cloud positional matching results.

(1) The nearest distance principle is used to obtain the relationship between the template point set $M = \{q_i\}$ $(i = 1, 2, 3, \ldots, L)$ and the corresponding points in the target point set $S = \{p_j\}$ $(j = 1, 2, 3, \ldots, K)$ to obtain the set of one-to-one pairs of points $\{(q_1, p_1), (q_2, p_2), \ldots; (q_i, p_i) \ldots; (q_K, p_K)\}$. The set of target points $S_w$ is the set of template points $M$, and the connecting line indicates the pairing relationship between the points. $R$ and $T$ are the rotational and translational transformation matrices between the target and model point sets as shown in Figure 8.

(2) Find the new positions of the model point set $M$ and the target point set $S$, respectively, after the centre of gravity coordinates $C_S, C_M$, and $p_i$ of the point cloud has been transformed by the $R$ matrix to $p'_i$; let the average pairwise distance in the transformed matching by ICP be $E$; the purpose of the ICP iteration is to find $R$, such that $E$ is minimized.

$$C_S = \frac{1}{N} \sum_{i=1}^{N} p_i,$$

$$C_M = \frac{1}{N} \sum_{i=1}^{N} q_i,$$  \hspace{1cm} (3)

$$p'_i = C_M + R(p_i - C_S),$$

$$E = \frac{1}{N} \sum_{i=1}^{N} \|q_i - p'_i\|^2.$$  

5. Experimental Results and Analysis

The object of this test is a white packaging bottle as shown in Figure 9. The two-dimensional image of this target object has good edge characteristics. The image was first acquired by the left camera in the binocular vision system and edge segmented to be able to distinguish and label individual individuals. The result of edge processing of the image is shown in Figure 10.

After edge processing, the image often has broken edges as shown in Figure 9, which is detrimental to the segmentation of objects using the connected-domain method and must be effectively edge repaired.

Figure 9 shows the partial edge fracture after Canny treatment and partial fracture repair.

The approach used in this study is as follows: for edge images, a Gaussian template is used for filtering and a Gaussian value is calculated for the white pixels close to the black pixels, resulting in reduction in the gray value of the white pixels. The Gaussian template is applied to the edge image and then binarised, increasing the width of the edge lines and effectively filling in the local breaks caused by the edge operator calculation.

6. Conclusions

A binocular 3D measurement system based on gray code structured light and phase-shift structured light was established by improving the robot to reconstruct the scene of scattering workpieces. In order to correctly segment the scattered parts and match the 3D pose of the target workpiece, this study first uses the edge features of the scattered parts to segment the 2D image to obtain the image area of the target workpiece and then uses the principle of optical geometric imaging to take the segmented 2D image as in segmenting a 3D point cloud. The three-dimensional visual inspection results of the parts are used to control the six-axis robot to grab the parts, and the effect is good. It is very helpful to our future production, greatly shortens the working hours, and reduces the cost of the enterprise.

Data Availability

The datasets used to support this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Q. Zhang, “A novel industrial robot grasping method based on binocular vision,” International Journal of Educational Economy and Management, vol. 37, no. 24, p. 26, 2017.

[2] M. Kumar and D. P. Garg, “Intelligent multi-sensor fusion techniques in flexible manufacturing workcells,”vol. 6, pp. 5375–5380, in Proceedings of the 2004 American Control Conference, vol. 6, IEEE, Boston, MA, USA, July 2004.

[3] Z. Zhang, Y. Li, F. Wang et al., “A novel multi-sensor environmental perception method using low-rank representation and a particle filter for vehicle reversing safety,” Sensors, vol. 16, no. 6, p. 848, 2016.

[4] D. Liu and L. Xu, “Optimal control method of robot end position and orientation based on dynamic tracking measurement,” IOP Conference Series: Earth and Environmental Science, vol. 108, no. 5, Article ID 52057, 2018.

[5] M. Song and S. Xin, "Robot autonomous sorting system for intelligent logistics," in Proceedings of the 2021 IEEE International Conference on Computer Science, Electronic
[6] C. Chen, X. Wu, L. Han, and Y. Ou, "Butler robot," in Proceedings of the 2011 IEEE International Conference on Information and Automation, pp. 732–737, IEEE, Shenzhen, China, June 2011.

[7] B. Wen, B. Tao, and G. Li, "Research status and tendency of intelligent industrial robot," Journal of Physics: Conference Series, vol. 1302, no. 3, Article ID 32050, 2019.

[8] Y. Tang, M. Chen, C. Wang et al., "Recognition and localization methods for vision-based fruit picking robots: a review," Frontiers of Plant Science, vol. 11, p. 510, 2020.

[9] L. Hu, J. Ou, J. Huang, Y. Chen, and D. Cao, "A review of research on traffic conflicts based on intelligent vehicles," IEEE Access, vol. 8, Article ID 24471, 2020.

[10] T. Jiang, H. Cui, X. Cheng, and W. Tian, "A measurement method for robot peg-in-hole prealignment based on combined two-level visual sensors," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–12, 2021.

[11] Y. Li, K. Yuan, Q. Cai et al., "Evolutionary bionic manipulator method based on capsule network," Journal of Intelligent and Fuzzy Systems, vol. 37, no. 4, pp. 4677–4685, 2019.

[12] Q. Bai, S. Li, J. Yang, Q. Song, Z. Li, and X. Zhang, "Object detection recognition and robot grasping based on machine learning: a survey," IEEE Access, vol. 8, Article ID 181855, 2020.

[13] H. Li, X.-m. Zhang, L. Zeng, and Y.-j. Huang, "A monocular vision system for online pose measurement of a 3RRR planar parallel manipulator," Journal of Intelligent and Robotic Systems, vol. 92, no. 1, pp. 3–17, 2018.

[14] Z. Li, T. Yang, G. Li, J. Li, and Y. Zhang, "Geodetic coordinate calculation based on monocular vision on UAV platform," in Proceedings of the 2016 IEEE 13th International Conference on Signal Processing (ICSP), pp. 310–314, IEEE, Chengdu, China, November 2016.

[15] A. Schyja, A. Hypki, and B. Kuhlenkötter, "A modular and extensible framework for real and virtual bin-picking environments," in Proceedings of the 2012 IEEE International Conference on Robotics and Automation, pp. 5246–5251, IEEE, Saint Paul, MN, USA, May 2012.

[16] M. Kumar and D. P. Garg, "Three-dimensional occupancy grids with the use of vision and proximity sensors in a robotic workcell," in Proceedings of the ASME International Mechanical Engineering Congress and Exposition, vol. 47063, pp. 1029–1036, Anaheim, CA, USA, November 2004.

[17] H. Zeng and J. Luo, "Construction of multi-modal perception model of communicative robot in non-structural cyber physical system environment based on optimized BT-SVM model," Computer Communications, vol. 181, pp. 182–191, 2022.

[18] J. Martinez-Gomez, A. Fernandez-Caballero, I. Garcia-Varea, L. Rodriguez, and C. Romero-Gonzalez, "A taxonomy of vision systems for ground mobile robots," International Journal of Advanced Robotic Systems, vol. 11, no. 7, p. 111, 2014.

[19] H. Yang, P. Jiang, and F. Wang, "Multi-view-based pose estimation and its applications on intelligent manufacturing," Sensors, vol. 20, no. 18, p. 5072, 2020.

[20] Z. Zhang, Y. Liu, and G. Lu, "Development of a movable service robot with double working arms for the elderly and the disabled," in Proceedings of the 2009 International Conference on Electronic Computer Technology, pp. 636–640, IEEE, Macau, China, February 2009.