EXTRACTION METHOD OF POSITION AND POSTURE INFORMATION OF ROBOT ARM PICKING UP TARGET BASED ON RGB-D DATA

by

Qi TENG[a], Jianjun YI[b*], Xiaoming ZHU[b], and Yajun ZHANG[b]

[a] Department of Computer Science and Engineering,
East China University of Science and Technology, Shanghai, China
[b] Department of Mechanical Engineering, East China University of Science and Technology,
Shanghai, China

Original scientific paper
https://doi.org/10.2298/TSCI190625012T

There is a big error in the traditional method to extract the position and attitude information of the robot. In the process of obtaining the target attitude, a method of extracting the target attitude information of robot arm based on RGB-D data is proposed. The position and attitude of the manipulator target are acquired by depth image processing, and the detected target position is sent to the manipulator control node, and the feature points of the manipulator are extracted. The 3-D mapping is carried out on the acquired RGB image, and the depth and RGB values of feature points, as well as position and attitude information are calculated by using the Gauss mixture model. Finally, the target is extracted by combining the covariance matrix of feature points. The experimental results show that the co-ordinate error and angle error of the robot arm extracted by this method are small. The maximum extraction error is only 28%, which is much lower than the traditional method, which shows that the proposed method is more applicable.

Key words: *RGB-D data, manipulator arm, pose information, dynamic target detection*

Introduction

Robot arm is the most widely used automatic mechanical device in the field of robotics technology [1]. It can be seen in the fields of industrial manufacturing, medical treatment, entertainment services, military, semiconductor manufacturing and space exploration [2]. Despite their different shapes, they all have a common feature, that is, they can accept instructions and accurately locate a point in 3-D (or 2-D) space for operation [3]. At present, there have been relevant research scholars, and some research results have been achieved.

Funes-Mora and Odobez [4] proposed a method to detect objects and their postures using RGB-D camera, which can provide real-time depth and color information. Model modeling process for each pose of the manipulator target grabbing by CAD model, and then finishes the initial estimation of the part pose by searching in 3-D space. Finally, ICP algorithm is used to achieve target grabbing. Li and Li [5] method used the defined shape descriptor to match the target and the model to complete the target attitude estimation. Ling *et al*. [6] method obtained the use of their point cloud information objects from the CAD model and used this information

* Corresponding author, e-mail: rcswkj1688@163.com

as a model base. Then, the model was used to estimate the attitude of the manipulator arm by iterating the nearest point (ICP) algorithm. Others have designed a series of point-to-point geometric edges of feature objects and used new features to obtain more accurate posture results. Although the previous methods can effectively recognize the pose of the pickup target, they depend on the precise background segmentation and the object database of random pose, there are some problems such as large error of target location information and poor robustness.

In order to solve the problem of the traditional aforementioned methods, a method based on RGB-D data is proposed to extract the position and attitude information of the robot arm.

Depth imaging process

The depth image obtained by RGB-D sensor can be understood as the distance from the observer's position to himself. That is to say, depth image is the image information that reflects the target distance detected by the sensor as the reference distance in the detection environment. So the depth image can also be understood as the distance image between the target and the sensor [7-9]. It is a practical method proposed by many scholars in China and abroad to map the depth image of the same scene into the RGB image space. By building a RGB-Depth system, the intrinsic relationship between the RGB image and the depth image of the same scene can be obtained [10].

In order to facilitate the later data transformation, it is necessary to quantify the point cloud data obtained by the sensor. The quantization process is co-ordinate transformation. The point cloud data based on the sensor co-ordinates are transformed into point cloud data in the world co-ordinates through calculation. This process is the basis and prerequisite for all data processing. The co-ordinate transformation under a single sensor is also a lot of engineering applications. The work foundation of multi-sensor fusion is that the co-ordinate system of point cloud data acquired by multi-sensor equipment is based on its own co-ordinate system [11]. There is no uniformity of data between devices. If we want to process all point cloud data, the first step is to put all point cloud data co-ordinates acquired by all devices into the world co-ordinate system, so that we can capture them from multiple angles. he quantization process of single sensor point cloud data is.

– Obtain the angle of view: The previous illustration is to obtain the angle range of the Xtion Pro live sensor in the horizontal direction: α is 57°. The vertical viewing angle β is 43°.
– Unit conversion ratio:

$$P_x = \frac{x}{z} = 2\tan\left(\frac{\alpha}{z}\right) \tag{1}$$

$$P_y = \frac{y}{z} = 2\tan\left(\frac{\beta}{z}\right) \tag{2}$$

In the formula, the proportions along the X-axis and along the Y-axis are expressed by P_x and P_y, respectively. The z represents the flexibility of robotic arms.

– Get the actual x and y values:

$$x_i = \left(\frac{x_{i0}}{n_x} - 0.5\right) z_{i0} p_x \tag{3}$$

The equation indicates that the right side of the depth map is the positive direction of the X-axis, while the left side of the depth map is the negative direction of the X-axis.

$$y_i = \left(0.5 - \frac{y_{i0}}{n_x}\right) z_{i0} p_y \tag{4}$$

The upper expression indicates that the upper edge of the depth map is the positive direction of the Y-axis, and the lower edge of the depth map is the negative direction of the Y-axis [12].

In the previous formula, the co-ordinate value of the midpoint of the actual space co-ordinate is (x_i, y_i), the co-ordinate value of the corresponding point in the imaging space is (x_{i0}, y_{i0}), and the resolution coefficients of RGB camera are n_x and n_y. For example, the resolution of RGB camera is 640 × 480, then the corresponding n_x is 640, n_y is 480, and z_{i0} is the depth value of the space point.

Extraction of position and posture information of target picked by manipulator arm

Firstly, the feature points of RGB image are extracted and mapped into 3-D space to obtain the 3-D co-ordinates of the feature points relative to the camera co-ordinate system, while preserving the RGB color information corresponding to the feature points.

Feature point extraction of robot arm picking target

When a new RGB-D data $F(x)$ arrives, the feature points of RGB image are extracted first. The commonly used feature extraction and description methods include scale invariant feature transformation (SIFT), SURF (accelerated robust feature), ORB (oriented FAST and rotated BRIEF) and Shi-Tomasi [13, 14]. Because SIFT and SURF methods are slow to extract feature points, while ORB is fast but not robust, Shi-Tomasi method is used to extract feature points. To some extent, this method takes both speed and robustness into account [15]. Because the extraction of descriptors and matching affect the real-time performance of the algorithm, only feature points need to be extracted without calculating their corresponding descriptors, as shown in fig. 1.

After extracting the pose feature points of the target image, the 2-D co-ordinates (u, v) and the corresponding color information (r, g, b) of each feature point in the image can be obtained, and then the 2-D co-ordinates can be mapped into 3-D space.

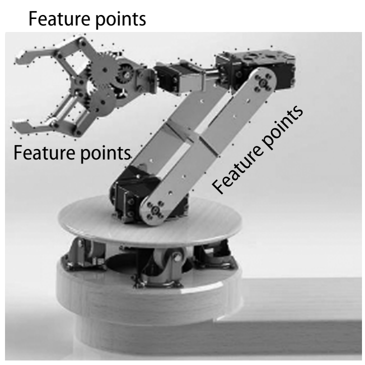

Figure 1. Extraction of feature points of position and posture for picking up target by manipulator arm

The 3-D mapping of characteristic points

In RGB image, the manipulator picks up the feature point (u, v) of the target's posture and position, whose depth value is z. The 3-D position of the point in the camera co-ordinate system is:

$$x^* = \frac{z}{f}\left(u - c_x\right) \tag{5}$$

$$y^* = \frac{z}{f}\left(v - c_y\right) \tag{6}$$

$$z^* = z \tag{7}$$

Among them, f is the focal length of RGB camera on Kinect sensor, while (c_x, c_y) is the center of the image. If the image resolution is assumed to be 640×480, then c_x and c_y are 320 and 240, respectively. According to the above formulas, the 3-D co-ordinate $[x^*, y^*, z^*]$ of $[u, v, z]$ relative to the camera co-ordinate system can be calculated. However, the position of μ_z is uncertain. Some studies consider that the depth measurement value is a Gaussian random variable with the mean value of μ_z, and the formula for calculating the standard deviation σ_z is:

$$\sigma_z = 1.45 \cdot 10^{-3} \mu_z^2 \tag{8}$$

Because the extracted feature points are usually located at the edge of the manipulator or where the color changes are obvious, as shown in fig. 1, the depth and color of the feature points are prone to jump [16, 17]. Therefore, only using the pixel (u, v) to get the object's 3-D co-ordinates and color will lead to larger errors in the depth and color measurement of feature points, which will lead to larger errors in x^* and y^*. At the same time, the location (u, v) of feature points also has some errors [18]. Through the analysis, the depth and RGB values of the feature points are calculated by using the Gauss mixture model.

Gauss mixture model

Assuming that there is a pixel error in the position of feature point (u, v), namely $\sigma_u = \sigma_v = 1$, nine pixels of 3 × 3 windows around feature point (u, v) are selected to calculate the depth value and variance of feature point. Assuming that the depth z_i of each pixel is a Gaussian distribution with the mean value of μ_{zi} and the variance of σ_{zi}^2, then according to the Gaussian mixture model [19], the mean μ_z and variance σ_z^2 of the depth of feature points are shown as follows:

$$\mu_z = \sum_{i=1}^{9} \omega_i \mu_{zi} \tag{9}$$

$$\sigma_z^2 = \sum_{i=1}^{9} \omega_i \left(\sigma_{zi}^2 + \mu_{zi}^2 \right) - \mu_z^2 \tag{10}$$

Among them, the weight at the feature point is 1/4, the weight at the top and bottom of the feature point is 1/8, and the weight at the oblique diagonal point is 1/16.

Similar to depth values, r, g and b at feature points are weighted sum of 9 pixels. Assuming that the color standard deviation σ_c is a constant, the r channel is taken as an example [18].

$$\mu_r = \sum_{i=1}^{9} \omega_i \mu_{ri} \tag{11}$$

$$\sigma_r^2 = \sum_{i=1}^{9} \omega_i \left(\sigma_c^2 + \mu_{ri}^2 \right) - \mu_r^2 \tag{12}$$

At this point, the 3-D co-ordinates and color values of the feature points, as well as the depth variance and color variance, are obtained.

Covariance matrix of characteristic points

By substituting the μ_z and (u, v) obtained from the previous formulas into the eqs. (5)-(7), the 3-D co-ordinate $\mu_{x,y,z} = [\mu_x, \mu_y, \mu_z]$ of the feature points after passing through the Gaussian mixture model can be obtained. Similarly, the $\mu_{r,g,b} = [\mu_r, \mu_g, \mu_b]$ can be obtained [19]. If the feature point is $\mu = [\mu_{x,y,z}, \mu_{r,g,b}]^T$, then the covariance matrix Σ is:

$$\begin{cases} \sum = \begin{bmatrix} \sum_{x,y,z} & 0 \\ 0 & \sum_{r,g,b} \end{bmatrix} \\ \sum_{x,y,z} = \begin{bmatrix} \sigma_x^2 & \sigma_{x,y} & \sigma_{x,z} \\ \sigma_{y,x} & \sigma_y^2 & \sigma_{y,z} \\ \sigma_{z,x} & \sigma_{z,y} & \sigma_z^2 \end{bmatrix} \\ \sum_{r,g,b} = \begin{bmatrix} \sigma_r^2 & & \\ & \sigma_r^2 & \\ & & \sigma_r^2 \end{bmatrix} \end{cases} \tag{13}$$

By synthesizing the previous formulas, it can be concluded that:

$$\begin{cases} \sigma_x^2 = \dfrac{\sigma_z^2 \left(\mu_u - c_x\right)^2 + \sigma_u^2 \mu_z^2}{f^2} \\ \sigma_y^2 = \dfrac{\sigma_z^2 \left(\mu_v - c_y\right)^2 + \sigma_v^2 \mu_z^2}{f^2} \\ \sigma_{x,y} = \sigma_{y,x} = \dfrac{\sigma_z^2 \left(\mu_u - c_x\right)\left(\mu_v - c_y\right)}{f^2} \\ \sigma_{x,z} = \sigma_{z,x} = \dfrac{\sigma_z^2 \left(\mu_u - c_x\right)}{f} \\ \sigma_{y,z} = \sigma_{z,y} = \dfrac{\sigma_z^2 \left(\mu_v - c_y\right)}{f} \end{cases} \tag{14}$$

By calculating, the position and attitude information $p = [x^*, y^*, z^*, r, g, b]^T$ of the pickup target can be approximated to the multivariate Gauss distribution with the mean value of $\mu = [\mu_{x,y,z}, \mu_{r,g,b}]^T$ and the covariance matrix of Σ.

Results

The proposed method of extracting position and pose information of picking up target by manipulator based on RGB-D data is simulated and analyzed. In the experiment, the effective ranging of detection module is generally about 100 m. In order to compare the accuracy of position and attitude information extraction between the proposed method and the current method, two methods are used to simulate and analyze the accuracy of position and attitude information extraction in the range of 100 m. In order to optimize the accuracy of position and attitude information extraction, experiments are carried out for several times, and the average calculation is carried out.

Record the data of each extraction experiment, and get the comparison between the extraction result and the actual value of the position and attitude information of the pickup target of the manipulator arm, as shown in tabs. 1 and 2.

Table 1. Comparisons betwen co-ordinates of principal feature points of posture and actual values

Experiment	Data	Co-ordinates of main feature points of posture and posture		
		X [mm]	*Y* [mm]	*Z* [mm]
1	Extraction value	852.8	647.4	156.7
	Actual value	853.1	647.2	157.0
2	Extraction value	697.2	680.1	155.9
	Actual value	696.9	680.5	156.3
3	Extraction value	495.8	786.3	143.8
	Actual value	496.2	786.0	144.1
4	Extraction value	588.6	743.1	128.9
	Actual value	588.9	743.3	129.2
5	Extraction value	397.4	827.5	149.6.
	Actual value	397.8	827.1	150.0
6	Extraction value	673.9	721.4	138.5
	Actual value	674.3	721.2	138.9

Table 2. Comparison of posture angle and actual value

Experiment	Data	Pose angle		
		Trend angle [°]	Pitch angle [°]	Flip angle [°]
1	Extraction value	–28.7	1.5	1.8
	Actual value	–29.1	1.6	1.7
2	Extraction value	–11.6	0.7	2.4
	Actual value	–12.5	0.6	2.5
3	Extraction value	–25.9	1.8	1.8
	Actual value	–25.3	2.0	1.9
4	Extraction value	–17.9	0.9	3.6
	Actual value	–18.1	0.7	3.8
5	Extraction value	–27.6	1.5	2.5
	Actual value	–25.8	1.6	2.7
6	Extraction value	–15.9	2.3	4.1
	Actual value	–16.2	2.2	3.9

In order to further verify the effectiveness of the multi-extraction algorithm, the other literature [4-6] methods are compared with the information extraction error as an index, the comparison results are shown in fig. 2.

Discussion

The data of tabs. 1 and 2 show that the position information error of the main feature points is less than 0.5 mm, the attitude angle error is less than 1.6°, and the attitude control deviation is smaller in the process of picking up the target of the manipulator arm, which can meet the control demand of picking up the target of the manipulator arm.

According to the analysis of fig. 2, when different methods are used to extract the position and attitude information of the target, the extraction error presents an upward trend. Among them, the error of [4] method is the largest, and the error of [5] is the largest. Comparatively speaking, the information extraction error of this method is lower. The maximum error is only 28%, which is much lower than other methods. It shows that the results obtained by this method are more reliable.

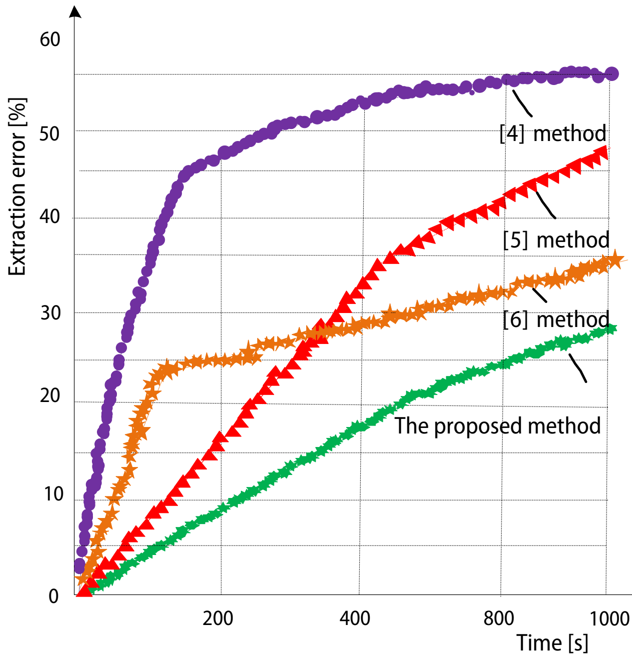

Figure 2. Error comparison of information extraction by different methods

Conclusions

In order to improve the efficiency and accuracy of the path planning and control of the manipulator, a method of extracting the position and attitude information of the manipulator based on RGB-D data is proposed. The position and attitude co-ordinates and angle data of the target are obtained by extracting the position and attitude information of the target picked up by the manipulator arm. Traditional RGB image processing is often accompanied by external information interference such as target shadows and sudden changes in light. Deep data detects targets by infrared, which is essentially distance information. This detection method is basically not affected by external light environment. For the control of the manipulator arm, the following problems need to be further studied. Path planning is one of the control core of intelligent manipulator. The current research results are basically based on theory. In the practical application process, the manipulator may appear vibration, stable pull and other situations, which is not conducive to the actual operation system. It is necessary to find effective methods to solve these problems in the follow-up study, so that the manipulator can perform tasks more efficiently.

References

[1] Choi, J., *et al.*, Twenty-One Degrees Of Freedom Model Based Hand Pose Tracking Using A Monocular RGB Camera, *Optical Engineering, 55* (2016), 1, 013101

[2] Chang, J. Y., Dr-Net: Denoising and Reconstruction Network for 3-D Human Pose Estimation from Monocular RGB Videos, *Electronics Letters, 54* (2018), 2, pp. 70-72
[3] Cui, Z., *et al.*, Simulation Study on the Trajectory Planning Problem of Redundant Robots (in Chinese), *Computer Simulation, 33* (2016), 3, pp. 268-273
[4] Funes-Mora, K. A., Odobez, J. M., Gaze Estimation in the 3D Space Using RGB-D Sensors, *International Journal of Computer Vision, 118* (2016), 2, pp. 194-216
[5] Li, J., Li, S., Gaze Estimation from Color Image Based on the Eye Model with Known Head Pose, *IEEE Transactions on Human-Machine Systems, 46* (2016), 3, pp. 1-10
[6] Ling, S., *et al.*, Performance Evaluation of Deep Feature Learning for RGB-D Image/Video Classification, *Information Sciences, 385-386* (2017), Apr., pp. 266-283
[7] Li, C. F., *et al.,* Realizing the Transformation of Martial Arts Education from the Existence of Posture to the Existence of Reality (in Chinese), *Journal of Guangzhou Sport University 37* (2017), 4, pp. 113-115
[8] Pandey, P. K., Jaboob, S. S. A., A Finite Difference Method for a Numerical Solution of Elliptic Boundary Value Problems, *Applied Mathematics & Nonlinear Sciences, 3* (2018), 1, pp. 311-320
[9] Brzezinski, D. W., Review of Numerical Methods for NumILPT with Computational Accuracy Assessment for Fractional Calculus, *Applied Mathematics & Nonlinear Sciences, 3* (2018), 2, pp. 487-502
[10] Li, X. Y., *et al.*, Attitude Control Algorithm for Simultaneous Control of Thruster Attitude and Orbit (in Chinese), *Journal of Jilin University (Science Edition), 54* (2016), 3, pp. 603-608
[11] Liu, H. B., *et al.*, Research on Information Interoperability Technology for Electric Vehicle Charging and Replacing Facilities (in Chinese), *Chinese Journal of Power Sources, 40* (2016), 12, pp. 2449-2452
[12] Hu, J. H., *et al.,* Spatial Information Cloud Service Platform Architecture and Application Research (in Chinese), *Journal of China Academy of Electronics, 11* (2016), 1, pp. 51-58
[13] Massarelli, C., *et al.,* Detection of Asbestos-Containing Materials in Agro-Ecosystem by the Use of Airborne Hyperspectral Casi-1500 Sensor Including the Limited Use of Two UAVS Equipped with RGB Cameras, *International Journal of Remote Sensing, 38* (2017), 8-10, pp. 2135-2149
[14] Ni, T., *et al.,* Real-Time Pose Control System for Robotic Arm Based on Kinect Dynamic Gesture Recognition, *Journal of Agricultural Machinery, 44* (2017), 10, pp. 417-423
[15] Qiu, X., *et al.,* Implementation of Visual Odometer Based on RGB-D Camera (in Chinese), *Journal of Zhejiang University of Technology, 45* (2017), 6, pp. 634-638
[16] Ren, B. Y., *et al.,* A Target Recognition and 3D Pose Estimation Method in Non-Structural Environment (in Chinese), *Journal of Harbin Institute of Technology, 51* (2019), 1, pp. 38-44
[17] Sun, W. L., Qian, L., Research on Attitude Calculation Based on Mems and Kalman Filter (in Chinese), *Automation & Instrumentation, 33* (2018), 1, pp. 1-3
[18] Tan, D. J., *et al.,* Real-Time Accurate 3D Head Tracking and Pose Estimation with Consumer RGB-D Cameras, *International Journal of Computer Vision, 126* (2017), 2-4, pp. 158-183
[19] Theobalt, C., Vnect: Real-Time 3D Human Pose Estimation with a Single RGB Camera, *Acm Transactions on Graphics, 36* (2017), 4, 44

Paper submitted: June 25, 2019
Paper revised: August 18, 2019
Paper accepted: August 19, 2019