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

Qi TENG1, Jianjun YI2*, Xiaoming ZHU2, Yajun ZHANG2

1Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

2Department of Mechanical Engineering, East China University of Science and Technology, Shanghai 200237, China

*rcswkj1688@163.com

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. 3D dimensional 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 coordinate 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

1 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 three-dimensional (or two-dimensional) space for operation [3]. At present, there have been relevant research scholars, and some research results have been achieved.

Reference [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 three-dimensional space. Finally, ICP algorithm is used to achieve target grabbing. Reference [5] method used the defined shape descriptor to match the target and the model to complete the target attitude estimation. Reference [6] method obtained the use of their point cloud information objects from the CAD model and used this information 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 above 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 methods mentioned above, a method based on RGB-D data is proposed to extract the position and attitude information of the robot arm.

2 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 coordinate transformation. The point cloud data based on the sensor coordinates are transformed into point cloud data in the world coordinates through calculation. This process is the basis and prerequisite for all data processing. The coordinate transformation under a single sensor is also a lot of engineering applications. The work foundation of multi-sensor fusion is that the coordinate system of point cloud data acquired by multi-sensor equipment is based on its own coordinate 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 coordinates acquired by all devices into the world coordinate system, so that we can capture them from multiple angles. he quantization process of single sensor point cloud data is as follows:

(1) Obtain the angle of view: The above illustration is to obtain the angle range of the Xtion Pro live sensor in the horizontal direction: $\alpha$ is 57°. The vertical viewing angle $\beta$ is 43°.

(2) Unit conversion ratio:

$$p_\alpha = \frac{x}{z} = 2 \tan \left( \frac{\alpha}{2} \right) \quad (1)$$
In the formula, the proportions along the X axis and along the Y axis are expressed by $P_x$ and $P_y$, respectively. $z$ represents the flexibility of robotic arms.

(3) Get the actual $x$ and $y$ values:

$$x_i = \left(\frac{X_0}{n_x} - 0.5\right) \times z_0 \times p_i$$  \hspace{1cm} (3)

$$y_i = (0.5 - \frac{Y_0}{n_y}) \times z_0 \times p_i$$  \hspace{1cm} (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 above formula, the coordinate value of the midpoint of the actual space coordinate is $(x_i, y_i)$, the coordinate value of the corresponding point in the imaging space is $(x_{im}, y_{im})$, 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 $z_0$ is the depth value of the space point.

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

Firstly, the feature points of RGB image are extracted and mapped into three-dimensional space to obtain the three-dimensional coordinates of the feature points relative to the camera coordinate system, while preserving the RGB color information corresponding to the feature points.

#### 3.1 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 SIFT (scale invariant feature transformation), 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 two-dimensional coordinates \((u,v)\) and the corresponding color information \((r,g,b)\) of each feature point in the image can be obtained, and then the two-dimensional coordinates can be mapped into three-dimensional space.

### 3.2 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 coordinate system is as follows:

\[
x^* = \frac{z}{f} (u-c_x) \tag{5}
\]

\[
y^* = \frac{z}{f} (v-c_y) \tag{6}
\]

\[
z^* = z \tag{7}
\]

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 \(640 \times 480\), then \(c_x\) and \(c_y\) are 320 and 240, respectively. According to the above formulas, the 3-D coordinate \([x^*,y^*,z^*]\) of \([u,v,z]\) relative to the camera coordinate system can be calculated. However, the position of \(\mu_z\) is uncertain. Some studies consider that the depth measurement value is a Gaussian random variable with the mean value of \(\mu_z\), and the formula for calculating the standard deviation \(\sigma_z\) is as follows:

\[
\sigma_z = 1.45 \times 10^{-5} \mu_z \tag{8}
\]

Because the extracted feature points are usually located at the edge of the manipulator or where the color changes are obvious, as shown in the figure above, 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 three-dimensional coordinates 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 above
analysis, the depth and RGB values of the feature points are calculated by using the Gauss mixture model.

### 3.3 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 \(\mu_i\) and the variance of \(\sigma_i^2\), then according to the Gaussian mixture model [19], the mean \(\mu_i\) and variance \(\sigma_i^2\) of the depth of feature points are shown as follows:

\[
\mu_i = \sum_{j=1}^{9} a_i \mu_j \quad (9)
\]

\[
\sigma_i^2 = \sum_{j=1}^{9} a_i (\sigma_j^2 + \mu_j^2) - \mu_i^2 \quad (10)
\]

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 \(\sigma_r\) is a constant, the \(r\) channel is taken as an example [18].

\[
\mu_i = \sum_{j=1}^{9} a_i \mu_j \quad (11)
\]

\[
\sigma_i^2 = \sum_{j=1}^{9} a_i (\sigma_j^2 + \mu_j^2) - \mu_i^2 \quad (12)
\]

At this point, the three-dimensional coordinates and color values of the feature points, as well as the depth variance and color variance, are obtained.

### 3.4 Covariance matrix of characteristic points

By substituting the \(\mu_i\) and \((u,v)\) obtained from the above formulas into the formulas (5), (6), (7), the three-dimensional coordinate \(\mu_{x,y,z} = [\mu_x, \mu_y, \mu_z]\) 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 \(\mu = [\mu_{x,y,z}, \mu_{r,g,b}]\), then the covariance matrix \(\Sigma\) is:
By synthesizing the above formulas, it can be concluded that:

\[
\begin{align*}
\sigma_r^2 &= \frac{\sigma_r^2 (\mu_r - c_r)^2}{f^2} + \sigma_r^2 \\
\sigma_y^2 &= \frac{\sigma_y^2 (\mu_y - c_y)^2}{f^2} + \sigma_y^2 \\
\sigma_z^2 &= \frac{\sigma_z^2 (\mu_z - c_z)^2}{f^2} + \sigma_z^2 \\
\sigma_{x,y} &= \sigma_{y,x} = \frac{\sigma_y^2 (\mu_x - c_x)(\mu_x - c_y)}{f^2} \\
\sigma_{x,z} &= \sigma_{z,x} = \frac{\sigma_z^2 (\mu_x - c_x)}{f} \\
\sigma_{y,z} &= \sigma_{z,y} = \frac{\sigma_z^2 (\mu_y - c_y)}{f}
\end{align*}
\]

(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 \( \mu = [\mu_x, \mu_y, \mu_z, \mu_r, \mu_g, \mu_b]^T \) and the covariance matrix of \( \Sigma \).

4 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 Tables 1 and 2.

Table 1 Comparisons between coordinates of principal feature points of posture and actual values
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Data</th>
<th>Coordinates of main feature points of posture and posture</th>
<th>X/mm</th>
<th>Y/mm</th>
<th>Z/mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Extraction Value</td>
<td>852.8</td>
<td>647.4</td>
<td>156.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>853.1</td>
<td>647.2</td>
<td>157.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Extraction Value</td>
<td>697.2</td>
<td>680.1</td>
<td>155.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>696.9</td>
<td>680.5</td>
<td>156.3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Extraction Value</td>
<td>495.8</td>
<td>786.3</td>
<td>143.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>496.2</td>
<td>786.0</td>
<td>144.1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Extraction Value</td>
<td>588.6</td>
<td>743.1</td>
<td>128.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>588.9</td>
<td>743.3</td>
<td>129.2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Extraction Value</td>
<td>397.4</td>
<td>827.5</td>
<td>149.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>397.8</td>
<td>827.1</td>
<td>150.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Extraction Value</td>
<td>673.9</td>
<td>721.4</td>
<td>138.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>674.3</td>
<td>721.2</td>
<td>138.9</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Comparison of posture angle and actual value

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Data</th>
<th>Pose Angle</th>
<th>Trend Angle (°)</th>
<th>Pitch Angle (°)</th>
<th>Flip Angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Extraction Value</td>
<td>-28.7</td>
<td>1.5</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>-29.1</td>
<td>1.6</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Extraction Value</td>
<td>-11.6</td>
<td>0.7</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>-12.5</td>
<td>0.6</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Extraction Value</td>
<td>-25.9</td>
<td>1.8</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>-25.5</td>
<td>2.0</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Extraction Value</td>
<td>-17.9</td>
<td>0.9</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>-18.1</td>
<td>0.7</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Extraction Value</td>
<td>-27.6</td>
<td>1.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>-25.8</td>
<td>1.6</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Extraction Value</td>
<td>-15.9</td>
<td>2.3</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actual Value</td>
<td>-16.2</td>
<td>2.2</td>
<td>3.9</td>
<td></td>
</tr>
</tbody>
</table>

In order to further verify the effectiveness of the multi-extraction algorithm, the other literature [4], literature [5] and literature [6] methods are compared with the information extraction error as an index, the comparison results are shown in Fig. 2.

![Fig. 2 Error comparison of information extraction by different methods](image-url)
5 Discussion

The data of Table 1 and Table 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 reference [4] method is the largest, and the error of reference [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.

6 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 coordinates 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


